

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using fw model

Run on: June 6, 2005, 09:17:48 ; Search time 166 Seconds  
(without alignments)  
1526.073 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSSVVERFIPVSGQNTNGF.....MIVFLTIALYKLFLKKYS 655

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.GeneSeq, 16Dec04: \*  
1: geneeqp1980g: \*  
2: geneeqp1990g: \*  
3: geneeqp2000g: \*  
4: geneeqp2001g: \*  
5: geneeqp2002g: \*  
6: geneeqp2003g: \*  
7: geneeqp2003b: \*  
8: geneeqp2004g: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3352	100.0	655	AAU80029	Aau80029 Human ABC
2	3352	100.0	663	AAV15221	Aay15221 Breast Ca
3	3346	99.8	655	AAAB60104	Aab60104 Human cta
4	3346	99.8	655	AAOI4781	Aaoi4781 Human BCR
5	3346	99.8	655	AAU80028	Aau80028 Human ABC
6	3346	99.8	655	ABRS8077	Abrs8077 Human ABC
7	3346	99.8	655	ADAI0917	Ada10917 Human CDN
8	3346	99.8	655	ADCS4182	Adc54182 Human bre
9	3346	99.8	655	ADG38394	Adg38394 Human w11
10	3346	99.8	655	AD157316	Ad157316 ATP-Bind1
11	3346	99.8	655	AD157315	Ad157315 ATP-Bind1
12	3346	99.8	655	AD157243	Ad157243 Human ATP
13	3346	99.8	655	AD157311	Ad157311 ATP-Bind1
14	3346	99.8	655	ADK67372	Adk67372 Human w11
15	3346	99.8	655	AD157314	Ad157314 ATP-Bind1
16	3346	99.7	655	ADG38390	Adg38390 Human BCR
17	3343	99.7	655	AD157310	Ad157310 ATP-Bind1
18	3343	99.7	655	ADG38388	Adg38388 Human BCR
19	3340	99.6	655	AD157312	Ad157312 ATP-Bind1
20	3339	99.6	655	AAOI4783	Aaoi4783 Human BCR
21	3338	99.6	655	ABBO7273	Abbo7273 Human BCR
22	3338	99.6	655	AD157313	Ad157313 ATP-Bind1
23	3331	99.4	655	AAV95365	Aay95365 ATP-Bind1
24	3331	99.4	655	AAU04348	Aau04348 Human BCR
25	3331	99.4	655	ABP52127	Abp52127 Homo sapi

26	3331	99.4	655	5	ABBO7270	Abbo7270 Human BCR
27	3331	99.4	655	6	ABU63376	Abu63376 Human mlt
28	3331	99.4	665	5	AAOI4782	Aaoi4782 Human BCR
29	3053.5	91.1	604	2	AAW73627	Aaw73627 Human sec
30	3053.5	91.1	604	5	ABP61858	Abp61858 Human pol
31	2927	87.3	623	8	ADJ27182	Adj27182 Murine BC
32	2757	82.2	657	5	ABBO7272	Abbo7272 Human pro
33	2325	62.4	456	4	AAAB93564	Aaab93564 Human pro
34	835.5	24.9	1049	8	ADN188679	Adn188679 Bacterial
35	812	24.2	687	4	ABBS9384	Abbs9384 Drosophila
36	809	24.1	751	8	ADN73449	Adn73449 Thale cre
37	808	24.1	687	3	AAV78981	Aay78981 Silkworm
38	795.5	23.7	676	8	ADN21146	Adn21146 Bacterial
39	783.5	23.4	625	3	AAAI8080	Aagi8080 Arabidops
40	783.5	23.4	632	3	AAI18079	Aai18079 Arabidops
41	783.5	22.4	648	3	AAI18078	Aai18078 Arabidops
42	746	22.3	649	3	AAQ43100	Aaq43100 Arabidops
43	746	22.3	653	3	AAQ43099	Aaq43099 Arabidops
44	739	22.0	678	8	ADN73697	Adn73697 Thale cre
45	727	21.7	832	4	ABBS9544	Abbs9544 Drosophila

## ALIGNMENTS

RESULT 1  
AAU80029 standard; protein; 655 AA.

AC AAU80029; DT 15-JUL-2002 (first entry)

XX Human ABCG2 mutant 482T.

DE Human ABCG2; transporter protein; anticancer drug tolerance;

KW Human; ABCG2; transporter protein; anticancer drug tolerance;

KM Indocarbazole; mutant; mutein.

XX Homo sapiens.

OS Synthetic.

FT Key Location/Qualifiers

FT Misc-difference 482 /note= "Wild type Arg substituted by Thr"

FN MO200228894-A1.

PD 11-APR-2002.

PF 18-SEP-2001; 2001MO-JP008112.

XX 03-OCT-2000; 2000JP-00303441.

XX (BANY ) BANYU PHARM CO LTD.

PI Komatani H, Hara Y, Kotani H, Nakagawa R;

XX WPI; 2002-352228/38.

DR N-PSDB; ABK49911.

PT ABCG2 gene encoding transporter protein capable of selectively

PT transporting indocarbazole compounds; useful in screening inhibitors and

PT anticancer agents for administration in chemotherapy.

PS Disclosure; Page 87-90; 98pp; Japanese.

XX The invention relates to an ABCG2 gene encoding a transporter protein

XX capable of imparting tolerance to an anticancer agent in mammals

XX comprising a fully defined sequence as given in the specification or an

XX amino acid sequence based on the sequence but with some amino acids

XX substituted, deleted or added. The gene and encoded protein are useful in

XX screening inhibitors and anticancer agents for administration in

XX chemotherapy with enhancement in sensitivity of cancer cell tolerance.

CC The gene relating to drug tolerance can be modified e.g. with the  
 CC transporter inhibitors, screened compounds, antibodies and antisense  
 CC nucleotides. The transporter is capable of selectively transporting  
 CC indocarbazole compounds extracellularly. The present sequence represents  
 CC the amino acid sequence of human ABCG2 mutant 482T

XX  
 XX  
 SQ Sequence 655 AA;

Query Match 100.0%; Score 3352; DB 5; Length 655;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSSNVEVPIPVSGQNTNGFPATASNDLKAFTGGAVALSFHNICYRVKLSGFLPCRKPYE 60  
 DB 1 MSSSNVEVPIPVSGQNTNGFPATASNDLKAFTGGAVALSFHNICYRVKLSGFLPCRKPYE 60  
 QY 61 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAARKDPSGLSDVILINGARPPANFKCN 120  
 DB 61 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAARKDPSGLSDVILINGARPPANFKCN 120  
 QY 121 SGVYVODDVVMGTLTVRENIQPSAALRLATTMTNHEKNERINRVIOELGLDVAADSKVGT 180  
 DB 121 SGVYVODDVVMGTLTVRENIQPSAALRLATTMTNHEKNERINRVIOELGLDVAADSKVGT 180  
 QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240  
 DB 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240  
 QY 241 SIHQPRYSIFKLPDLSLTLASGRLMFHPGPAQALGVESAGYHCEAVNNPADFFLDING 300  
 DB 241 SIHQPRYSIFKLPDLSLTLASGRLMFHPGPAQALGVESAGYHCEAVNNPADFFLDING 300  
 QY 301 DSTAVALNREDEPKATEIIEPSKODKPIEKLAIEIYNSFFYKETAELHQLSGEKKKK 360  
 DB 301 DSTAVALNREDEPKATEIIEPSKODKPIEKLAIEIYNSFFYKETAELHQLSGEKKKK 360  
 QY 361 ITVFEKISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGIVGAIYFGKNDK 420  
 DB 361 ITVFEKISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGIVGAIYFGKNDK 420  
 QY 421 TGIQNRAGVIFELITNOCSSVSAVELFVIEKKLFHEIYISGYRVSSTYFLGKLSDLP 480  
 DB 421 TGIQNRAGVIFELITNOCSSVSAVELFVIEKKLFHEIYISGYRVSSTYFLGKLSDLP 480  
 QY 481 MTMLPSIIFTCIYFPMGLKPKADAFVMMFTLMVAIYSSSMALIAAQSVVAATL 540  
 DB 481 MTMLPSIIFTCIYFPMGLKPKADAFVMMFTLMVAIYSSSMALIAAQSVVAATL 540  
 QY 541 MTICFVFMIFSGLVNLTITIASWLSWLYFSIPRYGFTALQNEFLGQNFPCGLNATGN 600  
 DB 541 MTICFVFMIFSGLVNLTITIASWLSWLYFSIPRYGFTALQNEFLGQNFPCGLNATGN 600  
 QY 601 NPGNVATCTGGEELVIVQGLDLSFWMGLMKHVALACMIVITLITAYLKLFKKYS 655  
 DB 601 NPGNVATCTGGEELVIVQGLDLSFWMGLMKHVALACMIVITLITAYLKLFKKYS 655

RESULT 2

AA15221 standard; protein; 663 AA.

AC AA15221;

DT 09-NOV-1999 (first entry)

DE Breast Cancer Resistance Protein (BCRP).

XX breast cancer; drug resistance; ATP-binding cassette; ABC;

KM xenobiotic transporter; chemotherapy; mitochondrion; doxorubicin;

XX breast cancer resistance protein; BCRP.

OS Homo sapiens.

XX

FH Key Location/Qualifiers  
 FT Domain 87..95  
 FT Domain /note="Walker A motif"  
 FT Domain 221..236  
 FT Domain /note="phosphopantetheine site"  
 FT Modified-site 345..347  
 FT Modified-site /note="Glycosylation site on N"  
 FT Region 405..422  
 FT Region /label=TM1  
 FT Modified-site /note="Transmembrane region"  
 FT Modified-site 425..427  
 FT Region /note="Glycosylation site on N"  
 FT Region 546..563  
 FT Modified-site /label=TM2  
 FT Modified-site 564..566  
 FT Modified-site /note="Glycosylation site on N"  
 FT Modified-site 604..606  
 FT Region /note="Glycosylation site on N"  
 FT Region 638..655  
 FT /label=TM3

W0940110-A1.

12-AUG-1999.

05-FEB-1999; 99WO-US002577.

05-FEB-1998; 98US-0073763P.

(UYMA-) UNIV MARYLAND BALTIMORE.

Ross DD, Doyle LA, Abruzzo L;

WPI: 1999-494273/41.

DR N-PSDB; AA206360.

PT New breast cancer resistance protein useful for production of antibodies

PS Claim 4; Fig 2a; 80pp; English.

XX The Breast Cancer Resistance Protein (BCRP) is an ATP-binding cassette  
 CC (ABC) transporter protein. It has a molecular mass of approximately 72.3  
 CC kilodaltons (kD) exclusive of any glycosylation. Expression of BCRP in  
 CC drug sensitive human cancer cells confers resistance to mitoxantrone,  
 CC doxorubicin, and daunorubicin, and reduces daunorubicin accumulation in  
 CC the cloned transfected cells. The protein is useful for producing  
 CC antibodies and antisense probes, which can be used to inhibit the  
 CC activity of BCRP, therefore enhancing a cancer patient's chemotherapy  
 CC treatment. The antibodies and probes overcomes the problems of breast  
 CC cancer resistance proteins to make chemotherapy treatment more effective

SQ Sequence 663 AA;

Query Match 100.0%; Score 3352; DB 2; Length 663;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSSNVEVPIPVSGQNTNGFPATASNDLKAFTGGAVALSFHNICYRVKLSGFLPCRKPYE 60

DB 9 MSSSNVEVPIPVSGQNTNGFPATASNDLKAFTGGAVALSFHNICYRVKLSGFLPCRKPYE 68

QY 61 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAARKDPSGLSDVILINGARPPANFKCN 120

DB 69 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAARKDPSGLSDVILINGARPPANFKCN 128

QY 121 SGVYVODDVVMGTLTVRENIQPSAALRLATTMTNHEKNERINRVIOELGLDVAADSKVGT 180

DB 129 SGVYVODDVVMGTLTVRENIQPSAALRLATTMTNHEKNERINRVIOELGLDVAADSKVGT 188

QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240

DB 189 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 248

QY 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFLDIING 300  
DB 249 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFLDIING 308  
QY 301 DSTAVALLNREDFPKATEIIEPSKODKPLIEKLAIEYVNSPFYKETAELHQSAGEKKKK 360  
DB 309 DSTAVALLNREDFPKATEIIEPSKODKPLIEKLAIEYVNSPFYKETAELHQSAGEKKKK 368  
QY 361 ITVFKEISYTTSFCHQLRWKSRSPKLLGNPOASIAQIIVTVVLGLVIGAIYFGLKXDS 420  
DB 369 ITVFKEISYTTSFCHQLRWKSRSPKLLGNPOASIAQIIVTVVLGLVIGAIYFGLKXDS 428  
QY 421 TGIQNRAGVLFPLTNNOCSSVSAVELFVVEKKLFHIEYISGYRVSSYFLGKLSDDL 480  
DB 429 TGIQNRAGVLFPLTNNOCSSVSAVELFVVEKKLFHIEYISGYRVSSYFLGKLSDDL 488  
QY 481 MTMLPSIIFTCIVFPMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540  
DB 489 MTMLPSIIFTCIVFPMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 548  
QY 541 MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQNFPCGINATGN 600  
DB 549 MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQNFPCGINATGN 608  
QY 601 NPCNVATCTGGEYLVKQIGIDLSPWGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655  
DB 609 NPCNVATCTGGEYLVKQIGIDLSPWGLMKHVALACMIVIFLTIAVLKLLFLKKYS 663

RESULT 3  
AAB60104  
ID AAB60104 standard; protein; 655 AA.  
XX AAB60104;  
XX 28-MAR-2001 (first entry)  
DT  
DE Human transport protein TPPT-24.  
XX  
KW Human; transport protein; TPPT; transport disorder; metabolic disorder;  
KW neurological disorder; cardiovascular disorder; reproductive disorder;  
KW immune disorder; cancer.  
OS Homo sapiens.  
XX  
PN WO200078953-A2.  
PD 28-DEC-2000.  
XX  
PF 16-JUN-2000; 2000WO-US016668.  
XX  
PR 17-JUN-1999; 99US-0139223P.  
PR 10-AUG-1999; 99US-0148177P.  
PR 18-AUG-1999; 99US-0149357P.  
PR 28-OCT-1999; 99US-0162287P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;  
PI Baughn MR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;  
XX  
DR WPT; 2001-041424/05.  
DR N-PSDB; AAF27724.  
XX  
XX Isolated polypeptide with a human transport protein sequence is useful  
PT for the diagnosis, prevention and treatment of disorders associated with  
PT the immune, reproductive and cardiovascular systems.  
XX  
PS Claim 2; Page 126-127; 165pp; English.  
XX  
CC The present invention provides the protein and coding sequences for 43  
CC novel human transport proteins (designated TPPTs). These can be used in

CC the diagnosis and treatment of transport, metabolic, neurological,  
CC reproductive, cardiovascular and immune disorders, and cell proliferative  
CC disorders such as cancer  
XX  
SQ Sequence 655 AA;  
Query Match 99.8%; Score 3346; DB 4; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MSSNVEVPIPVSGNTNGFPATASNDLKAFTEGVLSFHNICVYKLGKGLPCRPKE 60  
DB 1 MSSNVEVPIPVSGNTNGFPATASNDLKAFTEGVLSFHNICVYKLGKGLPCRPKE 60  
QY 61 KEILSININGIMKPGNALITPTGGKSSLDVLAARKPSGSLVINGARPPANFKCN 120  
DB 61 KEILSININGIMKPGNALITPTGGKSSLDVLAARKPSGSLVINGARPPANFKCN 120  
QY 121 SGYVQDDVNGTLTVRENLOFSALRLATTNHEKNERINRVIOELGLDKVADSKYGT 180  
DB 121 SGYVQDDVNGTLTVRENLOFSALRLATTNHEKNERINRVIOELGLDKVADSKYGT 180  
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQRTIIF 240  
DB 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQRTIIF 240  
QY 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFLDIING 300  
DB 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFLDIING 308  
QY 301 DSTAVALLNREDFPKATEIIEPSKODKPLIEKLAIEYVNSPFYKETAELHQSAGEKKKK 360  
DB 301 DSTAVALLNREDFPKATEIIEPSKODKPLIEKLAIEYVNSPFYKETAELHQSAGEKKKK 368  
QY 361 ITVFKEISYTTSFCHQLRWKSRSPKLLGNPOASIAQIIVTVVLGLVIGAIYFGLKXDS 420  
DB 361 ITVFKEISYTTSFCHQLRWKSRSPKLLGNPOASIAQIIVTVVLGLVIGAIYFGLKXDS 428  
QY 421 TGIQNRAGVLFPLTNNOCSSVSAVELFVVEKKLFHIEYISGYRVSSYFLGKLSDDL 480  
DB 421 TGIQNRAGVLFPLTNNOCSSVSAVELFVVEKKLFHIEYISGYRVSSYFLGKLSDDL 488  
QY 481 MTMLPSIIFTCIVFPMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540  
DB 481 MTMLPSIIFTCIVFPMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 548  
QY 541 MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQNFPCGINATGN 600  
DB 541 MTICVFPMIIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQNFPCGINATGN 608  
QY 601 NPCNVATCTGGEYLVKQIGIDLSPWGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655  
DB 601 NPCNVATCTGGEYLVKQIGIDLSPWGLMKHVALACMIVIFLTIAVLKLLFLKKYS 663

RESULT 4  
AA014781  
ID AA014781 standard; protein; 655 AA.  
XX AA014781;  
XX 28-JUN-2002 (first entry)  
DT  
DE Human BCRP protein.  
XX  
KW Human; BCRP protein; membrane penetrating region; cancer.  
XX  
OS Homo sapiens.  
XX  
PN JP2002065277-A.  
PD 05-MAR-2002.  
XX



DB 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEYVNSSFYKETKAEIHLQSGGEKXXX 360  
QY 361 ITVFEISYTTSPFCHQLRWVSKRSFKNLGNPOASIAOIIVTVVGLVIGALYFGLKXDS 420  
DB 361 ITVFEISYTTSPFCHQLRWVSKRSFKNLGNPOASIAOIIVTVVGLVIGALYFGLKXDS 420  
QY 421 TGIQRAGVLPFLTNOCFSSVSAVELFVEKKLFIHEISGYRVSSYFLGKLSDDL 480  
DB 421 TGIQRAGVLPFLTNOCFSSVSAVELFVEKKLFIHEISGYRVSSYFLGKLSDDL 480  
QY 481 MTMLPSIIFTCTIVFEMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540  
DB 481 MTMLPSIIFTCTIVFEMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540  
QY 541 MTICVFPMWIFSGLLVNLTTTASWLSWLOYSIPRYGFTALOHNEFLGNFCPGNATGN 600  
DB 541 MTICVFPMWIFSGLLVNLTTTASWLSWLOYSIPRYGFTALOHNEFLGNFCPGNATGN 600  
QY 601 NPCNVAATCTGEBYLVKQIGIDLSPMGLMKHVALACMIVIFLTITAYLKLFKXYS 655  
DB 601 NPCNVAATCTGEBYLVKQIGIDLSPMGLMKHVALACMIVIFLTITAYLKLFKXYS 655

RESULT 6  
ABR58077  
ID ABR58077 standard; protein; 655 AA.  
XX ABR58077;  
AC 15-OCT-2003 (first entry)  
XX DT  
XX DE Human ABCG2 protein.  
XX KW ABCG2; antidiabetic; cell therapy; diabetes mellitus;  
XX KM pancreatic stem cell; islets of langerhans; insulin.  
XX OS Homo sapiens.  
XX PN WO2003026584-A2.  
XX PD 03-APR-2003.  
XX PE 26-SEP-2002; 2002MO-US030700.  
XX PR 26-SEP-2001; 2001US-00963875.  
XX PR 11-APR-2002; 2002US-00120687.  
XX PR 02-MAY-2002; 2002US-00136891.  
XX PA (GENO) GEN HOSPITAL CORP.  
XX P1 Habener JF, Zulewski H, Thomas MK, Abraham EJ, Vallejo M;  
XX P1 Leech CA, Nolan AL, Lechner A;  
XX DR MPI; 2003-354625/33.  
XX DR N-PSDB; ACC80605.  
XX PT Treating a patient with diabetes mellitus by isolating a nestin- or ABCG2  
XX PT -positive pancreatic stem cell from a pancreatic islet of a donor and  
XX PT transferring the stem cell into the patient.  
XX PS Disclosure; Fig 18B; 107pp; English.

CC The invention relates to a method of treating a patient with diabetes  
CC mellitus by isolating a nestin- or ABCG2-positive pancreatic stem cell  
CC from a pancreatic islet of a donor, and transferring the stem cell into  
CC the patient whereby the stem cell differentiates into an insulin-  
CC producing cell. Alternatively, the nestin- or ABCG2-positive stem is  
CC induced into a pancreatic progenitor cell prior to isolation and  
CC transfer. This sequence corresponds to the human ABCG2 protein and the  
CC encoding gene is detected in the method of the invention. The method is  
CC useful for preparing a pharmaceutical composition for treating diabetes  
CC mellitus. The stem cells can be further characterised for correct gene

CC expression using the primers and probes ACC80607-ACC80671  
XX SQ Sequence 655 AA;  
Query Match 99.8%; Score 3346; DB 6; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MSSNVEVFIPVSOQNTNGFPATASNDLKAFEGAVLSPHNCYRVKLSGFLPCRPYE 60  
DB 1 MSSNVEVFIPVSOQNTNGFPATASNDLKAFEGAVLSPHNCYRVKLSGFLPCRPYE 60  
QY 61 KEILSININGIMKPGNALILPTGGKSSLDVLAARKDPSGSLGVLLINGAPRPANFKCN 120  
DB 61 KEILSININGIMKPGNALILPTGGKSSLDVLAARKDPSGSLGVLLINGAPRPANFKCN 120  
QY 121 SGYVQDDVWGTLTVRENLOPSAALRLATTMTNHEKNERIRRVIOELGLDQVADSKVGT 180  
DB 121 SGYVQDDVWGTLTVRENLOPSAALRLATTMTNHEKNERIRRVIOELGLDQVADSKVGT 180  
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSANAVLLLRMSKQGRITIF 240  
DB 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSANAVLLLRMSKQGRITIF 240  
QY 241 SIHQPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCAVNNPADFLDIING 300  
DB 241 SIHQPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCAVNNPADFLDIING 300  
QY 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEYVNSSFYKETKAEIHLQSGGEKXXX 360  
DB 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEYVNSSFYKETKAEIHLQSGGEKXXX 360  
QY 361 ITVFEISYTTSPFCHQLRWVSKRSFKNLGNPOASIAOIIVTVVGLVIGALYFGLKXDS 420  
DB 361 ITVFEISYTTSPFCHQLRWVSKRSFKNLGNPOASIAOIIVTVVGLVIGALYFGLKXDS 420  
QY 421 TGIQRAGVLPFLTNOCFSSVSAVELFVEKKLFIHEISGYRVSSYFLGKLSDDL 480  
DB 421 TGIQRAGVLPFLTNOCFSSVSAVELFVEKKLFIHEISGYRVSSYFLGKLSDDL 480  
QY 481 MTMLPSIIFTCTIVFEMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540  
DB 481 MTMLPSIIFTCTIVFEMGLKPKADAFVVMFTLMVAVSASSMALAIAAGSVSVATLL 540  
QY 541 MTICVFPMWIFSGLLVNLTTTASWLSWLOYSIPRYGFTALOHNEFLGNFCPGNATGN 600  
DB 541 MTICVFPMWIFSGLLVNLTTTASWLSWLOYSIPRYGFTALOHNEFLGNFCPGNATGN 600  
QY 601 NPCNVAATCTGEBYLVKQIGIDLSPMGLMKHVALACMIVIFLTITAYLKLFKXYS 655  
DB 601 NPCNVAATCTGEBYLVKQIGIDLSPMGLMKHVALACMIVIFLTITAYLKLFKXYS 655

RESULT 7  
ADA10917  
ID ADA10917 standard; protein; 655 AA.  
XX ADA10917;  
AC 06-NOV-2003 (first entry)  
XX DT  
XX DE Human CDNA differentially expressed in colon cancer #23 product.  
XX KW differential expression; colon cancer; cancer; human.  
XX OS Homo sapiens.  
XX PN US2002160382-A1.  
XX PD 31-OCT-2002.  
XX PF 11-OCT-2001; 2001US-00981353.  
XX

```

PR 11-OCT-2000; 2000US-0239841P.
XX (LASEK/) LASEK A W.
PA (JONE/) JONES D A.
XX
PI Lasek AW, Jones DA;
XX
DR WPI; 2003-265756/26.
DR N-PSDB; ADA10916.
XX
PT New combination comprising cDNAs that are differentially expressed in
PT colon disorder, useful for diagnosing, treating, staging or monitoring
PT treatment for colon cancers.
XX
XX Example 14; SEQ ID NO 35; 231bp; English.
XX
XX The invention relates to a combination comprising cDNAs that are
XX differentially expressed in colon disorder. The methods and compositions
XX of the present invention are useful for diagnosing, treating, staging or
XX monitoring treatment for colon cancer. They are also useful in high
XX throughput methods for using cDNAs to detect differential expression of
XX nucleic acids in a sample, screening molecules or compounds to identify a
XX ligand which specifically binds a cDNA and using a protein to screen
XX molecules or compounds to identify at least one ligand which specifically
XX binds the protein. The present sequence represents the amino acid
XX sequence of a human cDNA differentially expressed in colon cancer
XX protein.
XX
XX Sequence 655 AA:
XX
Query Match 99.8%; Score 3346; DB 6; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRYLKSGLPCRRPYE 60
DB 1 MSSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRYLKSGLPCRRPYE 60
QY 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGLSDVLINGAPRPANFKCN 120
DB 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGLSDVLINGAPRPANFKCN 120
QY 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGLSDVLINGAPRPANFKCN 120
DB 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGLSDVLINGAPRPANFKCN 120
QY 121 SGYVODDVVMGTLTYRENIQFSALRLATTTNHEKNERINRYIOELGLDKYADSKVGT 180
DB 121 SGYVODDVVMGTLTYRENIQFSALRLATTTNHEKNERINRYIOELGLDKYADSKVGT 180
QY 121 SGYVODDVVMGTLTYRENIQFSALRLATTTNHEKNERINRYIOELGLDKYADSKVGT 180
DB 121 SGYVODDVVMGTLTYRENIQFSALRLATTTNHEKNERINRYIOELGLDKYADSKVGT 180
QY 181 QFIRGVSGBGRKRTSICMELITDPSILFDEPTTGDSSTANAVLLLLKMSQGRITIF 240
DB 181 QFIRGVSGBGRKRTSICMELITDPSILFDEPTTGDSSTANAVLLLLKMSQGRITIF 240
QY 241 SIHQPRYSIFKLFDLSITLASGRLMFHPAQEALGYESAGYCEAVNPNADFFLDING 300
DB 241 SIHQPRYSIFKLFDLSITLASGRLMFHPAQEALGYESAGYCEAVNPNADFFLDING 300
QY 241 SIHQPRYSIFKLFDLSITLASGRLMFHPAQEALGYESAGYCEAVNPNADFFLDING 300
DB 241 SIHQPRYSIFKLFDLSITLASGRLMFHPAQEALGYESAGYCEAVNPNADFFLDING 300
QY 301 DSTAVALNREDFKATEIIPSPKODKPLIEKLAIEIYVNSFYETKAEHLQLSGGEKKK 360
DB 301 DSTAVALNREDFKATEIIPSPKODKPLIEKLAIEIYVNSFYETKAEHLQLSGGEKKK 360
QY 361 ITVFKESYTTSCHOURLRWYSKRSFKNLGNPQASTAQIIVTVVLGIVGAIYFGKLNDS 420
DB 361 ITVFKESYTTSCHOURLRWYSKRSFKNLGNPQASTAQIIVTVVLGIVGAIYFGKLNDS 420
QY 421 TGIQNBAGVLFLLTNQCFSSVAVELFVEVEKLFIEHYISGYRVSSYFLGKLSDLPL 480
DB 421 TGIQNBAGVLFLLTNQCFSSVAVELFVEVEKLFIEHYISGYRVSSYFLGKLSDLPL 480
QY 481 MTMLPSIIIFTCTIYFMLGLKPKADAFVMMFTLMVAVASASMAIAIAGQSVSVATLL 540
DB 481 MTMLPSIIIFTCTIYFMLGLKPKADAFVMMFTLMVAVASASMAIAIAGQSVSVATLL 540
QY 541 MTICFVFMMLFSGILVNLTTIASLWSLQYFESIPIRYGFTALQNEBFLGNFCGLNATGN 600
DB 541 MTICFVFMMLFSGILVNLTTIASLWSLQYFESIPIRYGFTALQNEBFLGNFCGLNATGN 600

```

```

QY 601 NPCNATCTGEBEYLKQIDLSPMGLMKHVALACMIVIFLTAYLKLFLKXYS 655
DB 601 NPCNATCTGEBEYLKQIDLSPMGLMKHVALACMIVIFLTAYLKLFLKXYS 655

RESULT 8
ADCS4182
ID ADCS4182 standard; protein; 655 AA.
XX
AC ADCS4182;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human breast cancer resistance protein (BCRP) amino acid sequence.
XX
XX Cancer cell; anti-cancer agent; steroid hormone; oestrogenic effect;
XX BCRP; breast cancer resistance protein; cytosolic; camptothecin;
XX mitoxantrone; 7-hydroxy staurosporine; adriamycin; cancer chemotherapy;
XX human.
XX
XX Homo sapiens.
XX
XX JP2003063989-A.
XX
XX 05-MAR-2003.
XX
XX 23-AUG-2001; 2001JP-00252953.
XX
XX 23-AUG-2001; 2001JP-00252953.
XX
XX (GANK-) ZH GAN KENKYUKAI.
XX
XX WPI; 2003-735321/70.
XX
XX N-PSDB; ADCS4181.
XX
PT Agent that overcomes resistance of cancer cell against anti-cancer agent,
PT comprises a steroid hormone, or a compound which exhibits antagonistic
PT activity against the hormone, with the cancer cell expressing BCRP gene.
XX
XX Example 1; SEQ ID NO 4; 15bp; Japanese.
XX
XX This invention relates to a novel agent which overcomes resistance of a
XX cancer cell against an anti-cancer agent (AA), comprising as an active
XX ingredient a steroid hormone, a compound having oestrogenic effect, or a
XX compound which exhibits antagonistic activity against the hormone, where
XX the cancer cell expresses the BCRP (breast cancer resistance protein)
XX gene. The agent of the invention may have cytostatic activity. The
XX invention is useful for overcoming resistance of a cancer against an anti
XX -cancer agent such as camptothecin, mitoxantrone, 7-hydroxy
XX staurosporine and adriamycin. The therapeutic effective anti-cancer agent
XX is recovered, due to the use of the agent of the invention. Also the
XX dosages of anti-cancer agent can be maintained easily, and adverse
XX effects of cancer chemotherapy can be suppressed. The present sequence is
XX that of the human BCRP protein which was used to develop the novel agent
XX of the invention.
XX
XX Sequence 655 AA:
XX
Query Match 99.8%; Score 3346; DB 7; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRYLKSGLPCRRPYE 60
DB 1 MSSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRYLKSGLPCRRPYE 60
QY 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGLSDVLINGAPRPANFKCN 120
DB 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGLSDVLINGAPRPANFKCN 120
QY 121 SGYVODDVVMGTLTYRENIQFSALRLATTTNHEKNERINRYIOELGLDKYADSKVGT 180
DB 121 SGYVODDVVMGTLTYRENIQFSALRLATTTNHEKNERINRYIOELGLDKYADSKVGT 180

```

```

Db      121 SGVYVODDVWGTLTVRENLQPSAALRLATTTWTHNEKNERINRVOELGLDKVADSKVGT 180
Qy      181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVILLLRKMSKQGRITIF 240
Db      181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVILLLRKMSKQGRITIF 240
Qy      241 SIHOPRYSIFKLFDSTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
Db      241 SIHOPRYSIFKLFDSTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
Qy      301 DSTAVALNREEDFPKATEIIEPSKODKPLIEKLAETVNSSFYKETVAELHQLSGGEKKK 360
Db      301 DSTAVALNREEDFPKATEIIEPSKODKPLIEKLAETVNSSFYKETVAELHQLSGGEKKK 360
Qy      361 ITVFKETSYTTSFCHQLRWKRSKSFKNLLGNPQASIAQIIVTVVGLVIGALYFGIKNDS 420
Db      361 ITVFKETSYTTSFCHQLRWKRSKSFKNLLGNPQASIAQIIVTVVGLVIGALYFGIKNDS 420
Qy      421 TGIQRAGVLPFLTNQCFSSVSAVELFVVEKKLFHEHISGYRVSSYFLGKLSDDL 480
Db      421 TGIQRAGVLPFLTNQCFSSVSAVELFVVEKKLFHEHISGYRVSSYFLGKLSDDL 480
Qy      481 MTMLPSIIFTCIVYFMLGLKPKADAFVMMFTLMVAVASASSMALAIAAGQSVSVATLL 540
Db      481 MTMLPSIIFTCIVYFMLGLKPKADAFVMMFTLMVAVASASSMALAIAAGQSVSVATLL 540
Qy      541 MTICVPMWIFSGGLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGONFCPGLNATGN 600
Db      541 MTICVPMWIFSGGLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGONFCPGLNATGN 600
Qy      601 NPCNVAATCTGGEYLVKQIDLSPMGLMKHVALACMIVIFLTIAVKKLFLKKYS 655
Db      601 NPCNVAATCTGGEYLVKQIDLSPMGLMKHVALACMIVIFLTIAVKKLFLKKYS 655

```

## RESULT 9

ADG38394  
ID ADG38394 standard; protein; 655 AA.

ADG38394;

26-FEB-2004 (first entry)

Human wild-type BCRP.

Anticancer agent; polymorphism; human; BCRP; cancer cell.

Homo sapiens.

JP2003199585-A.

15-JUL-2003.

21-MAY-2002; 2002JP-00145926.

24-OCT-2001; 2001JP-00325883.

(GANK-) ZH GAN KENKUKAI.

WPI; 2003-819597/77.

N-Psdb; ADG38395.

Evaluating sensitivity of test cell to anticancer agent involves

identifying gene polymorphism of BCRP.

Example 1; SEQ ID NO 7; 18bp; Japanese.

The present invention relates to a method for evaluating the sensitivity of a cell to an anticancer agent. The method involves identifying a gene polymorphism in the human BCRP gene (the polymorphism is undefined in the specification). The gene polymorphisms encode variant BCRP polypeptides designated as Q141K, V14M and Q126STOP. Identifying the gene polymorphism of BCRP of a test cell is useful for evaluating the expression grade of

CC the side effect at the time of administering an anticancer agent to the  
CC test cell and evaluating the resistance of the test cell to the  
CC anticancer agent. BCRP protein is useful in conveying an anticancer agent  
CC to cancer cell. The method is efficient in identifying a safer anticancer  
CC agent for treatment. The present sequence represents wild-type BCRP.

XX Sequence 655 AA;

Query Match 99.8%; Score 3346; DB 7; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1 MSSNVEVPIPSQNTNGFPATASNDLKAFFEGAVLSFHNICYRVKLSGFLPCRKPYE 60
Db      1 MSSNVEVPIPSQNTNGFPATASNDLKAFFEGAVLSFHNICYRVKLSGFLPCRKPYE 60
Qy      61 KEILSNINGIMKPGINAILGPTGGKSSLDVLAARKDPSGLSDVYLNGAPRPANFKCN 120
Db      61 KEILSNINGIMKPGINAILGPTGGKSSLDVLAARKDPSGLSDVYLNGAPRPANFKCN 120
Qy      121 SGVYVODDVWGTLTVRENLQPSAALRLATTTWTHNEKNERINRVOELGLDKVADSKVGT 180
Db      121 SGVYVODDVWGTLTVRENLQPSAALRLATTTWTHNEKNERINRVOELGLDKVADSKVGT 180
Qy      181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVILLLRKMSKQGRITIF 240
Db      181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVILLLRKMSKQGRITIF 240
Qy      241 SIHOPRYSIFKLFDSTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
Db      241 SIHOPRYSIFKLFDSTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
Qy      301 DSTAVALNREEDFPKATEIIEPSKODKPLIEKLAETVNSSFYKETVAELHQLSGGEKKK 360
Db      301 DSTAVALNREEDFPKATEIIEPSKODKPLIEKLAETVNSSFYKETVAELHQLSGGEKKK 360
Qy      361 ITVFKETSYTTSFCHQLRWKRSKSFKNLLGNPQASIAQIIVTVVGLVIGALYFGIKNDS 420
Db      361 ITVFKETSYTTSFCHQLRWKRSKSFKNLLGNPQASIAQIIVTVVGLVIGALYFGIKNDS 420
Qy      421 TGIQRAGVLPFLTNQCFSSVSAVELFVVEKKLFHEHISGYRVSSYFLGKLSDDL 480
Db      421 TGIQRAGVLPFLTNQCFSSVSAVELFVVEKKLFHEHISGYRVSSYFLGKLSDDL 480
Qy      481 MTMLPSIIFTCIVYFMLGLKPKADAFVMMFTLMVAVASASSMALAIAAGQSVSVATLL 540
Db      481 MTMLPSIIFTCIVYFMLGLKPKADAFVMMFTLMVAVASASSMALAIAAGQSVSVATLL 540
Qy      541 MTICVPMWIFSGGLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGONFCPGLNATGN 600
Db      541 MTICVPMWIFSGGLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGONFCPGLNATGN 600
Qy      601 NPCNVAATCTGGEYLVKQIDLSPMGLMKHVALACMIVIFLTIAVKKLFLKKYS 655
Db      601 NPCNVAATCTGGEYLVKQIDLSPMGLMKHVALACMIVIFLTIAVKKLFLKKYS 655

```

## RESULT 10

ADIS7316  
ID ADIS7316 standard; protein; 655 AA.

ADIS7316;

22-APR-2004 (first entry)

ATP-binding cassette transporter ABCG2 D590Y mutant.

drug transport capability; polymorphism; ABCG2; polymorphic mutation;

drug sensitivity; anti-cancer drug; cancer therapy;

cancer cell detection; indolocarbazole compound; human;

ABC transporter superfamily;

ATP-binding cassette transporter superfamily; mutant; mutenin.



OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 590 /note= "Wild type Asp substituted by Tyr"  
 XX  
 PN WO2003107249-A1.  
 XX  
 PD 24-DEC-2003.  
 XX  
 PF 13-JUN-2003; 2003WO-JP007534.  
 XX  
 PR 17-JUN-2002; 2002JP-00175806.  
 XX  
 PA (BANY ) BANYU PHARM CO LTD.  
 XX  
 PI Kotani H, Mizuara S;  
 DR WPI; 2004-156349/15.  
 XX  
 PT Predicting drug transport capability of mammalian cell by collecting  
 sample from mammal, determining polymorphism of nucleotide sequence of  
 ABCG2 gene or polymorphism of amino acid sequence of ABCG2 polypeptide.  
 XX  
 PS Example 1; Page; 76pp; English.  
 XX  
 CC The invention describes a method of predicting a drug transport  
 capability of a mammalian cell involving collecting a sample from a  
 mammal, determining a polymorphism of the nucleotide sequence of ABCG2  
 gene or a polymorphism of the amino acid sequence of ABCG2 polypeptide.  
 CC The method is useful for predicting drug transport capability of a  
 mammalian cell. Polynucleotides comprising single nucleotide  
 CC polymorphisms or polynucleotides comprising polymorphic mutations of the  
 ABCG2 protein are useful as diagnostic agent for diagnosing drug  
 CC sensitivity which involves analyzing a biological sample from a subject  
 CC and determining the presence or absence of the polynucleotides or  
 CC polypeptides, where the subject having the polynucleotide and/or the  
 CC polypeptide is suggested to be sensitive to the indolocarbazole compound.  
 CC A transformed cell comprising an ABCG2 protein mutant is useful for  
 CC measuring drug transport capability. By predicting drug transport  
 CC capability of a mammalian cell, sensitivity of a patient to various drugs  
 CC such as anti-cancer drugs can be diagnosed and an indicator for the  
 CC therapy can be obtained. As a result of selecting an anti-cancer drug in  
 CC cancer therapy and, particularly, detecting a cancer cell(s) which is  
 CC highly sensitive to indolocarbazole compounds, it is now possible to  
 CC selectively apply the compounds for the therapy. In addition, the optimum  
 CC dose of the indolocarbazole compounds in the cancer therapy is found and,  
 CC at the same time, side effect of the compounds is reduced whereby a  
 CC highly effective method of using the indolocarbazole compounds is  
 CC provided. This is the amino acid sequence of a human ABC transporter  
 CC superfamily (ATP-binding cassette transporter superfamily) protein ABCG2  
 CC mutant. Note: This sequence does not appear in the specification but has  
 CC been created using information given in the claims of the invention.  
 CC  
 XX  
 SQ Sequence 655 AA;  
 Query Match 99.8%; Score 3346; DB 8; Length 655;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGDDSTANAVLLLRKMSKQGRITIF 240  
 DB 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGDDSTANAVLLLRKMSKQGRITIF 240  
 QY 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGTFESAGHYCEAVNNPADFFLDIING 300  
 DB 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGTFESAGHYCEAVNNPADFFLDIING 300  
 QY 301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNVSFYKETAEHLHQLSGEKKK 360  
 DB 301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNVSFYKETAEHLHQLSGEKKK 360  
 QY 361 ITVFKESIYTTSPCHQLRWMSKRSFKNLLGNPQASIAQIIVTVVLGIVGAIYFGKND 420  
 DB 361 ITVFKESIYTTSPCHQLRWMSKRSFKNLLGNPQASIAQIIVTVVLGIVGAIYFGKND 420  
 QY 421 TGIQNRAGVLFLLTNOCFSSVSAGVLLFVVEKKLFHEIYISGYRVSSTYGLKLLDLP 480  
 DB 421 TGIQNRAGVLFLLTNOCFSSVSAGVLLFVVEKKLFHEIYISGYRVSSTYGLKLLDLP 480  
 QY 481 MTMLPSIIFTCIYFNLGLKPKADAFVVMFTLMVAYASASWALAIAGQSVSVATLL 540  
 DB 481 MTMLPSIIFTCIYFNLGLKPKADAFVVMFTLMVAYASASWALAIAGQSVSVATLL 540  
 QY 541 MTICFVFMNIFSGILVNLTTIASWLSMLOYFSIPRYGFTALQNEFLQNFCEGLNATGN 600  
 DB 541 MTICFVFMNIFSGILVNLTTIASWLSMLOYFSIPRYGFTALQNEFLQNFCEGLNATGN 600  
 QY 601 NPGNYATCGEELVVGQIDLSPPWGLMKXHVALLACMIVFLTAYLKLFLKYS 655  
 DB 601 NPGNYATCGEELVVGQIDLSPPWGLMKXHVALLACMIVFLTAYLKLFLKYS 655  
 RESULT 11  
 ADI57315  
 ID ADI57315 standard; protein; 655 AA.  
 AC ADI57315;  
 XX  
 DT 22-APR-2004 (first entry)  
 XX  
 DE ATP-binding cassette transporter ABCG2 R482T mutant.  
 XX  
 KW drug transport capability; polymorphism; ABCG2; polymorphic mutation;  
 KW drug sensitivity; anti-cancer drug; cancer therapy;  
 KW cancer cell detection; indolocarbazole compound; human;  
 KW ABC transporter superfamily;  
 KW ATP-binding cassette transporter superfamily; mutant; mutcin.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Misc-difference 482 /note= "Wild type Arg substituted by Thr"  
 FT  
 PN WO2003107249-A1.  
 PD 24-DEC-2003.  
 PF 13-JUN-2003; 2003WO-JP007534.  
 XX  
 PR 17-JUN-2002; 2002JP-00175806.  
 XX  
 PA (BANY ) BANYU PHARM CO LTD.  
 XX  
 PI Kotani H, Mizuara S;  
 DR WPI; 2004-156349/15.  
 XX  
 PT Predicting drug transport capability of mammalian cell by collecting  
 sample from mammal, determining polymorphism of nucleotide sequence of  
 ABCG2 gene or polymorphism of amino acid sequence of ABCG2 polypeptide.



Example 1; Page: 76pp; English.

The invention describes a method of predicting a drug transport capability of a mammalian cell involving collecting a sample from a mammal, determining a polymorphism of the nucleotide sequence of ABCG2 gene or a polymorphism of the amino acid sequence of ABCG2 polypeptide. The method is useful for predicting drug transport capability of a mammalian cell. Polynucleotides comprising single nucleotide polymorphisms or polypeptides comprising polymorphic mutations of the ABCG2 protein are useful as diagnostic agent for diagnosing drug sensitivity which involves analyzing a biological sample from a subject and determining the presence or absence of the polynucleotides or polypeptides, where the subject having the polynucleotide and/or the polypeptide is suggested to be sensitive to the indolocarbazole compound. A transformed cell comprising an ABCG2 protein mutant is useful for measuring drug transport capability. By predicting drug transport capability of a mammalian cell, sensitivity of a patient to various drugs such as anti-cancer drugs can be diagnosed and an indicator for the therapy can be obtained. As a result of selecting an anti-cancer drug in cancer therapy and, particularly, detecting a cancer cell(s) which is highly sensitive to indolocarbazole compounds, it is now possible to selectively apply the compounds for the therapy. In addition, the optimum dose of the indolocarbazole compounds in the cancer therapy is found and, at the same time, side effect of the compounds is reduced whereby a highly effective method of using the indolocarbazole compounds is provided. This is the amino acid sequence of a human ABC transporter superfamily (ATP-binding cassette transporter superfamily) protein ABCG2 mutant. Note: This sequence does not appear in the specification but has been created using information given in the claims of the invention.

Sequence 655 AA;

Query Match 99.8%; Score 3346; DB 8; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MSSNVEVPIPSQNTNGFPATASNDLKAFTEGAVLSFPHNICYVVKLSGSLPCKPVE 60
DB 1 MSSNVEVPIPSQNTNGFPATASNDLKAFTEGAVLSFPHNICYVVKLSGSLPCKPVE 60
QY 61 KEILNININGIMKPGNALIGPTGGGKSSLDVLAARKDPSGSGVLLINGAPRPANFKCN 120
DB 61 KEILNININGIMKPGNALIGPTGGGKSSLDVLAARKDPSGSGVLLINGAPRPANFKCN 120
QY 121 SGYVVDVVMGTLTVRENLOFSALRLATTMTNHEKNERINRVIOELGLDQVADSKVGT 180
DB 121 SGYVVDVVMGTLTVRENLOFSALRLATTMTNHEKNERINRVIOELGLDQVADSKVGT 180
QY 181 QPIRVSQGERKRTSIGMELITDPSILFDEPTTGIDSTANAVALLLKRMKQKQRTIIF 240
DB 181 QPIRVSQGERKRTSIGMELITDPSILFDEPTTGIDSTANAVALLLKRMKQKQRTIIF 240
QY 241 SIHQRYSTIFKLPDSITLLASGRLMFHGPAQALGFESAGVHCEAYNNPADFLDIING 300
DB 241 SIHQRYSTIFKLPDSITLLASGRLMFHGPAQALGFESAGVHCEAYNNPADFLDIING 300
QY 301 DSTAVANLRDEDFKATEIIEPSKODKPLIEKLAETIVNSSFYKETKAEHLQSGEKKKK 360
DB 301 DSTAVANLRDEDFKATEIIEPSKODKPLIEKLAETIVNSSFYKETKAEHLQSGEKKKK 360
QY 361 ITVFEISTYTSFCHQLRWVSKSRFNLLGNPOASIAQIIVVVGIVGAIYFGLKND 420
DB 361 ITVFEISTYTSFCHQLRWVSKSRFNLLGNPOASIAQIIVVVGIVGAIYFGLKND 420
QY 421 TGIQRAGVLPFLTNQCPSSVSAVELFVEKKLFHEIYSYXRYSSYFLKSLDLLP 480
DB 421 TGIQRAGVLPFLTNQCPSSVSAVELFVEKKLFHEIYSYXRYSSYFLKSLDLLP 480
QY 481 MTMLPSIIFTCIYVFMGLKPRADAFVVMFTLMVAVYSASSMALAIAAGQSVSATLL 540
DB 481 MTMLPSIIFTCIYVFMGLKPRADAFVVMFTLMVAVYSASSMALAIAAGQSVSATLL 540

```

```

QY 541 MTICFVPMWIPSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNFLGNCPCGINATCN 600
DB 541 MTICFVPMWIPSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNFLGNCPCGINATCN 600
QY 601 NPCNATCTGSEYLYKQIGIDSPWGLMKRNHVALACMIYIFLITAYLKLLFLKYS 655
DB 601 NPCNATCTGSEYLYKQIGIDSPWGLMKRNHVALACMIYIFLITAYLKLLFLKYS 655

RESULT 12
AD157243
ID AD157243 standard; protein; 655 AA.
AC AD157243;
XX 22-APR-2004 (first entry)
DT
DE Human ATP-binding cassette transporter ABCG2.
XX drug transport capability; polymorphism; ABCG2; polymorphic mutation;
XX cancer cell detection; indolocarbazole compound; human;
XX ABC transporter superfamily;
XX ATP-binding cassette transporter superfamily.
OS Homo sapiens.
XX MO2003107249-A1.
XX 24-DEC-2003.
XX 13-JUN-2003; 2003MO-JP007534.
XX 17-JUN-2002; 2002JP-00175806.
XX (BANY ) BANYU PHARM CO LTD.
XX Kotani H, Mizutani S;
XX WPI, 2004-156349/15.
XX N-PSDB; AD157242.
XX
XX Predicting drug transport capability of mammalian cell by collecting
XX sample from mammal, determining polymorphism of nucleotide sequence of
XX ABCG2 gene or polymorphism of amino acid sequence of ABCG2 polypeptide.
XX Claim 16; SEQ ID NO 2; 76pp; English.
XX
XX The invention describes a method of predicting a drug transport
XX capability of a mammalian cell involving collecting a sample from a
XX mammal, determining a polymorphism of the nucleotide sequence of ABCG2
XX gene or a polymorphism of the amino acid sequence of ABCG2 polypeptide.
XX The method is useful for predicting drug transport capability of a
XX mammalian cell. Polynucleotides comprising single nucleotide
XX polymorphisms or polypeptides comprising polymorphic mutations of the
XX ABCG2 protein are useful as diagnostic agent for diagnosing drug
XX sensitivity which involves analyzing a biological sample from a subject
XX and determining the presence or absence of the polynucleotides or
XX polypeptides, where the subject having the polynucleotide and/or the
XX polypeptide is suggested to be sensitive to the indolocarbazole compound.
XX A transformed cell comprising an ABCG2 protein mutant is useful for
XX measuring drug transport capability. By predicting drug transport
XX capability of a mammalian cell, sensitivity of a patient to various drugs
XX such as anti-cancer drugs can be diagnosed and an indicator for the
XX cancer therapy can be obtained. As a result of selecting an anti-cancer drug in
XX cancer therapy and, particularly, detecting a cancer cell(s) which is
XX highly sensitive to indolocarbazole compounds, it is now possible to
XX selectively apply the compounds for the therapy. In addition, the optimum
XX dose of the indolocarbazole compounds in the cancer therapy is found and,
XX at the same time, side effect of the compounds is reduced whereby a
XX highly effective method of using the indolocarbazole compounds is
XX provided. This is the amino acid sequence of human ABC transporter
XX superfamily (ATP-binding cassette transporter superfamily) protein ABCG2.

```

XX SQ Sequence 655 AA;  
 Query Match 99.8%; Score 3346; DB 8; Length 655;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSSNVEVFIPIVQSGNTNGFPATASNDLKAFTGAVLSFHNICYRYKLSGFLPCRKYE 60  
 |||||  
 DB 1 MSSSNVEVFIPIVQSGNTNGFPATASNDLKAFTGAVLSFHNICYRYKLSGFLPCRKYE 60  
 |||||

QY 61 KEILSNINGIMKGLNAILGPTGGKSLDVLAAKDPGSLGVDYLINGAPPANFKCN 120  
 |||||  
 DB 61 KEILSNINGIMKGLNAILGPTGGKSLDVLAAKDPGSLGVDYLINGAPPANFKCN 120  
 |||||

QY 121 SGVYVDDVVMGTLTYRENLOFSAALRLATTMNHKNERINVIQELGDKYADSVGT 180  
 |||||  
 DB 121 SGVYVDDVVMGTLTYRENLOFSAALRLATTMNHKNERINVIQELGDKYADSVGT 180  
 |||||

QY 181 QFIRGVSGERKRTSIGMELITDPSILFDEPTTGSDSTANAVLLLLKMSKQGRITIF 240  
 |||||  
 DB 181 QFIRGVSGERKRTSIGMELITDPSILFDEPTTGSDSTANAVLLLLKMSKQGRITIF 240  
 |||||

QY 241 SIHQPRYSIFKLFDSLTLASGRLMFHPAQEALGYFESAGYHCEAYNNPADPFLDIING 300  
 |||||  
 DB 241 SIHQPRYSIFKLFDSLTLASGRLMFHPAQEALGYFESAGYHCEAYNNPADPFLDIING 300  
 |||||

QY 301 DSTAVALNREEDKATEIIEPSKODKPLEKLEIYVNSFEYKETAELHQLSGGEKKK 360  
 |||||  
 DB 301 DSTAVALNREEDKATEIIEPSKODKPLEKLEIYVNSFEYKETAELHQLSGGEKKK 360  
 |||||

QY 361 ITVFKEISYTTSEFCHOLRWVSKSPKMLGNPOASIAQIIVTVVLGIVGAIYFGKKNDS 420  
 |||||  
 DB 361 ITVFKEISYTTSEFCHOLRWVSKSPKMLGNPOASIAQIIVTVVLGIVGAIYFGKKNDS 420  
 |||||

QY 421 TGIQNRAGVLFELTTNOCSSVASVBLEFVVEKLFHEIYSGYRVSIFLGKLSDDL 480  
 |||||  
 DB 421 TGIQNRAGVLFELTTNOCSSVASVBLEFVVEKLFHEIYSGYRVSIFLGKLSDDL 480  
 |||||

QY 481 MTMLPSIIFTCIYFMYLGLKPKADAFVMMFTLMWVYSSASSMALIAAGQSVVATLL 540  
 |||||  
 DB 481 MTMLPSIIFTCIYFMYLGLKPKADAFVMMFTLMWVYSSASSMALIAAGQSVVATLL 540  
 |||||

QY 541 MTICFVPMMFSGILVNLTTIASMLSWLYFSIPRYGFTALQHNPELQNFCEGLNATGN 600  
 |||||  
 DB 541 MTICFVPMMFSGILVNLTTIASMLSWLYFSIPRYGFTALQHNPELQNFCEGLNATGN 600  
 |||||

QY 601 NPCNVATCTGGEYLVKQIDLSPWGLMKHVALACMIVIFLTTAYLKLFLKXYS 655  
 |||||  
 DB 601 NPCNVATCTGGEYLVKQIDLSPWGLMKHVALACMIVIFLTTAYLKLFLKXYS 655  
 |||||

RESULT 13  
 ADIS7311  
 ID ADIS7311 standard; protein; 655 AA.  
 XX  
 XX ADIS7311;  
 DT 22-APR-2004 (first entry)  
 XX  
 DE ATP-binding cassette transporter ABCG2 Q141K mutant.  
 XX  
 XX drug transport capability; polymorphism; ABCG2; polymorphic mutation;  
 KM drug sensitivity; anti-cancer drug; cancer therapy;  
 KM cancer cell detection; indolocarbazole compound; human;  
 KM ABC transporter superfamily;  
 KM ATP-binding cassette transporter superfamily; mutant; muten.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 141

FT /note= "wild type Gln substituted by Lys"  
 XX WO2003107249-A1.  
 XX  
 XX 24-DEC-2003.  
 XX  
 XX 13-JUN-2003; 2003WO-UP007534.  
 XX  
 XX 17-JUN-2002; 2002JP-00175806.  
 XX  
 XX (BANY ) BANYU PHARM CO LTD.  
 XX  
 XX Kotani H, Mizunari S;  
 XX  
 XX WPI; 2004-156349/15.  
 XX  
 PT Predicting drug transport capability of mammalian cell by collecting  
 PT sample from mammal, determining polymorphism of nucleotide sequence of  
 PT ABCG2 gene or polymorphism of amino acid sequence of ABCG2 polypeptide.  
 XX  
 XX Claim 6; Page; 76pp; English.

CC The invention describes a method of predicting a drug transport  
 CC capability of a mammalian cell involving collecting a sample from a  
 CC mammal, determining a polymorphism of the nucleotide sequence of ABCG2  
 CC gene or a polymorphism of the amino acid sequence of ABCG2 polypeptide.  
 CC The method is useful for predicting drug transport capability of a  
 CC mammalian cell. Polynucleotides comprising single nucleotide  
 CC polymorphisms or polypeptides comprising polymorphic mutations of the  
 CC ABCG2 protein are useful as diagnostic agent for diagnosing drug  
 CC sensitivity which involves analyzing a biological sample from a subject  
 CC and determining the presence or absence of the polynucleotides or  
 CC polypeptides, where the subject having the polynucleotide and/or the  
 CC polypeptide is suggested to be sensitive to the indolocarbazole compound.  
 CC A transformed cell comprising an ABCG2 protein mutant is useful for  
 CC measuring drug transport capability. By predicting drug transport  
 CC capability of a mammalian cell, sensitivity of a patient to various drugs  
 CC such as anti-cancer drugs can be diagnosed and an indicator for the  
 CC therapy can be obtained. As a result of selecting a cancer cell(s) which is  
 CC highly sensitive to indolocarbazole compounds, it is now possible to  
 CC selectively apply the compounds for the therapy. In addition, the optimum  
 CC dose of the indolocarbazole compounds in the cancer therapy is found and,  
 CC at the same time, side effect of the compounds is reduced whereby a  
 CC highly effective method of using the indolocarbazole compounds is  
 CC provided. This is the amino acid sequence of a human ABC transporter  
 CC superfamily (ATP-binding cassette transporter superfamily) protein ABCG2  
 CC mutant. Note: This sequence does not appear in the specification but has  
 CC been created using information given in the claims of the invention.  
 XX

SQ Sequence 655 AA;  
 Query Match 99.8%; Score 3346; DB 8; Length 655;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSSNVEVFIPIVQSGNTNGFPATASNDLKAFTGAVLSFHNICYRYKLSGFLPCRKYE 60  
 |||||  
 DB 1 MSSSNVEVFIPIVQSGNTNGFPATASNDLKAFTGAVLSFHNICYRYKLSGFLPCRKYE 60  
 |||||

QY 61 KEILSNINGIMKGLNAILGPTGGKSLDVLAAKDPGSLGVDYLINGAPPANFKCN 120  
 |||||  
 DB 61 KEILSNINGIMKGLNAILGPTGGKSLDVLAAKDPGSLGVDYLINGAPPANFKCN 120  
 |||||

QY 121 SGVYVDDVVMGTLTYRENLOFSAALRLATTMNHKNERINVIQELGDKYADSVGT 180  
 |||||  
 DB 121 SGVYVDDVVMGTLTYRENLOFSAALRLATTMNHKNERINVIQELGDKYADSVGT 180  
 |||||

QY 181 QFIRGVSGERKRTSIGMELITDPSILFDEPTTGSDSTANAVLLLLKMSKQGRITIF 240  
 |||||  
 DB 181 QFIRGVSGERKRTSIGMELITDPSILFDEPTTGSDSTANAVLLLLKMSKQGRITIF 240  
 |||||

QY 241 SIHQPRYSIFKLFDSLTLASGRLMFHPAQEALGYFESAGYHCEAYNNPADPFLDIING 300  
 |||||

```

Db      241 SIHQPRYSIFKLPDSTLLASGRLMFHGPAQBALGYFESAGYHCEAYNNPADFLDIING 300
Qy      301 DSTAVALNREDFPKATEIIEPSKODKPLIEKLAETIYVNSFYKETAELHOLSGGEKKK 360
Db      301 DSTAVALNREDFPKATEIIEPSKODKPLIEKLAETIYVNSFYKETAELHOLSGGEKKK 360
Qy      361 ITVFKEISYTTSPFCHQLRWVSKRSPKNLLGNPOASIAQIIVTVVLGVIGAIYFGLKND 420
Db      361 ITVFKEISYTTSPFCHQLRWVSKRSPKNLLGNPOASIAQIIVTVVLGVIGAIYFGLKND 420
Qy      421 TGIQNRAGVLPFLTTNOCFSSVSAVELFVEKKLFIHEYISGYRVSSYFLGKLSDDL 480
Db      421 TGIQNRAGVLPFLTTNOCFSSVSAVELFVEKKLFIHEYISGYRVSSYFLGKLSDDL 480
Qy      481 MTMLPSIIFTCIYVFMGLKPKADAFVVMFTLMVAVASASMALAIAAGSVSVATLL 540
Db      481 MTMLPSIIFTCIYVFMGLKPKADAFVVMFTLMVAVASASMALAIAAGSVSVATLL 540
Qy      541 MTICVFPMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELGONFCPLGNATGN 600
Db      541 MTICVFPMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELGONFCPLGNATGN 600
Qy      601 NPCNVATCTGEBEYLVKQIDLSPMGLMKHVALACMIVIFLTIAYLKLFLEKYS 655
Db      601 NPCNVATCTGEBEYLVKQIDLSPMGLMKHVALACMIVIFLTIAYLKLFLEKYS 655

```

## RESULT 14

ADK67372  
ID ADK67372 standard; protein; 655 AA.

AC ADK67372;

DT 18-NOV-2004 (first entry)

DE Human wild-type ABCG2 (ATP-binding cassette gene) protein.

XX drug absorption; ABCG2; ATP-binding cassette gene; human; wild-type;

KM chromosome 4q22.

XX Homo sapiens.

OS JP2004016042-A.

PN 22-JAN-2004.

XX 13-JUN-2002; 2002JP-00172759.

PF 13-JUN-2002; 2002JP-00172759.

XX 13-JUN-2002; 2002JP-00172759.

PA (KOKU-) KOKURITSU IYAKUHIN SHOKUHIN EISEI KENKYU.

XX (IYAK-) IYAKUHIN FUKUSAYO HIGAI KYUSAI KENKYU SH.

DR WPI; 2004-113852/12.

XX N-PSDB; ADK67371.

XX Novel ABCG2 polynucleotide having a mutation at a specific position,

PT associated with ABCG2 protein.

XX Claim 1; SEQ ID NO 2; 53pp; Japanese.

CC The invention relates to a novel polynucleotide having a mutation in the  
CC codon encoding a glutamine residue present at the 126 position of a 655  
CC amino acid sequence. The polynucleotide of the invention may be useful  
CC for the estimation or diagnosis of a condition which is associated with  
CC abnormal drug absorption and in which the ABCG2 (ATP-binding cassette  
CC gene) protein is associated. The current sequence is that of the human  
CC wild-type ABCG2 protein of the invention which is encoded by DNA located  
CC at chromosome 4q22.

XX Sequence 655 AA;

Query Match 99.8%; Score 3346; DB 8; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1 MSSNVEVPIPVSGNTNGPAPATASNDLKAFTEGAVLSPHNI CYRVKLSGFLPCRKY 60
Db      1 MSSNVEVPIPVSGNTNGPAPATASNDLKAFTEGAVLSPHNI CYRVKLSGFLPCRKY 60
Qy      61 KEILSNINGIMKPGNALITPGGKSLDLVLAARKDPSGLSGVLNGARPPANFKN 120
Db      61 KEILSNINGIMKPGNALITPGGKSLDLVLAARKDPSGLSGVLNGARPPANFKN 120
Qy      121 SGYVQDDVVMGTLVRENLOFSALRLATTMTHEKNERIRNVQELGLDVAASKYCT 180
Db      121 SGYVQDDVVMGTLVRENLOFSALRLATTMTHEKNERIRNVQELGLDVAASKYCT 180
Qy      181 QPIRVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLKRMKSGKRTIIF 240
Db      181 QPIRVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLKRMKSGKRTIIF 240
Qy      241 SIHQPRYSIFKLPDSTLLASGRLMFHGPAQBALGYFESAGYHCEAYNNPADFLDIING 300
Db      241 SIHQPRYSIFKLPDSTLLASGRLMFHGPAQBALGYFESAGYHCEAYNNPADFLDIING 300
Qy      301 DSTAVALNREDFPKATEIIEPSKODKPLIEKLAETIYVNSFYKETAELHOLSGGEKKK 360
Db      301 DSTAVALNREDFPKATEIIEPSKODKPLIEKLAETIYVNSFYKETAELHOLSGGEKKK 360
Qy      361 ITVFKEISYTTSPFCHQLRWVSKRSPKNLLGNPOASIAQIIVTVVLGVIGAIYFGLKND 420
Db      361 ITVFKEISYTTSPFCHQLRWVSKRSPKNLLGNPOASIAQIIVTVVLGVIGAIYFGLKND 420
Qy      421 TGIQNRAGVLPFLTTNOCFSSVSAVELFVEKKLFIHEYISGYRVSSYFLGKLSDDL 480
Db      421 TGIQNRAGVLPFLTTNOCFSSVSAVELFVEKKLFIHEYISGYRVSSYFLGKLSDDL 480
Qy      481 MTMLPSIIFTCIYVFMGLKPKADAFVVMFTLMVAVASASMALAIAAGSVSVATLL 540
Db      481 MTMLPSIIFTCIYVFMGLKPKADAFVVMFTLMVAVASASMALAIAAGSVSVATLL 540
Qy      541 MTICVFPMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELGONFCPLGNATGN 600
Db      541 MTICVFPMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELGONFCPLGNATGN 600
Qy      601 NPCNVATCTGEBEYLVKQIDLSPMGLMKHVALACMIVIFLTIAYLKLFLEKYS 655
Db      601 NPCNVATCTGEBEYLVKQIDLSPMGLMKHVALACMIVIFLTIAYLKLFLEKYS 655

```

## RESULT 15

AD157314  
ID AD157314 standard; protein; 655 AA.

AC AD157314;

DT 22-APR-2004 (first entry)

DE ATP-binding cassette transporter ABCG2 R482G mutant.

XX drug transport capability; polymorphism; ABCG2; polymorphic mutation;

KM drug sensitivity; anti-cancer drug; cancer therapy;

XX cancer cell detection; indolocarbazole compound; human;

OS ABC transporter superfamily;

XX ATP-binding cassette transporter superfamily; mutant; mutein.

OS Homo sapiens.

XX Synthetic.

XX Key

FT Misc-difference 482

Location/Qualifiers  
/note= "Wild type Arg substituted by Gly"



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2005, 09:30:34 / Search time 43 Seconds  
(without alignments)  
1137.096 Million cell updates/sec

Title: US-09-961-086-1

Perfect score: 3352  
Sequence: 1 MSSSVVEVIFPVSGQNTNGF.....MIVFLITAIYKLPLKKYS 655

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patentg\_A: \*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/6C\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/6D\_COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3352	100.0	655	US-09-245-808-1	Sequence 1, Appl1
2	3331	99.4	655	US-09-767-594-1	Sequence 1, Appl1
3	835.5	24.9	1049	US-09-538-092-72	Sequence 72, Appl1
4	706.5	21.1	674	US-09-538-092-1125	Sequence 1125, Ap
5	693.5	20.7	652	US-09-989-981A-2	Sequence 2, Appl1
6	682.5	20.4	651	US-09-989-981A-6	Sequence 6, Appl1
7	657.5	19.6	672	US-09-989-981A-4	Sequence 4, Appl1
8	640.5	19.1	673	US-09-989-981A-8	Sequence 8, Appl1
9	567	16.9	617	US-09-614-912-138	Sequence 138, App
10	537	16.0	1296	US-09-614-912-140	Sequence 140, App
11	516	15.4	740	US-09-248-796A-20212	Sequence 20212, A
12	514.5	15.3	518	US-09-248-796A-20772	Sequence 20772, A
13	488.5	14.6	1411	US-09-538-092-413	Sequence 413, App
14	483.5	14.4	520	US-09-270-767-43468	Sequence 43468, A
15	463	13.8	300	US-09-248-796A-200770	Sequence 200770, A
16	461.5	13.8	694	US-09-248-796A-20210	Sequence 20210, A
17	461	13.8	539	US-09-614-912-144	Sequence 144, App
18	430	12.8	611	US-09-248-796A-20762	Sequence 20762, A
19	425.5	12.7	602	US-09-248-796A-20057	Sequence 20057, A
20	371	11.1	286	US-09-248-796A-200214	Sequence 200214, A
21	370.5	11.1	626	US-09-248-796A-200211	Sequence 200211, A
22	333	9.9	266	US-09-248-796A-20072	Sequence 20072, A
23	298.5	8.9	358	US-09-248-796A-20070	Sequence 20070, A
24	254	7.6	198	US-10-162-012-36	Sequence 36, Appl
25	253.5	7.6	290	US-09-270-767-41454	Sequence 41454, A
26	250	7.5	389	US-09-248-796A-16734	Sequence 16734, A
27	246	7.3	270	US-09-134-001C-5561	Sequence 5561, Ap

28	240.5	7.2	385	4	US-09-270-767-45978	Sequence 45978, A
29	240.5	7.2	653	4	US-09-543-681A-5411	Sequence 5411, Ap
30	239	7.1	247	4	US-09-603-208A-144	Sequence 144, App
31	238.5	7.1	1684	3	US-08-665-252-25	Sequence 25, Appl
32	238.5	7.1	1684	3	US-08-762-500-25	Sequence 25, Appl
33	238.5	7.1	1704	3	US-08-762-500-75	Sequence 75, Appl
34	238.5	7.1	1704	4	US-09-032-438C-120	Sequence 120, App
35	238.5	7.1	1766	4	US-09-949-016-10796	Sequence 10796, A
36	233.5	7.0	247	4	US-09-902-540-11438	Sequence 11438, A
37	232.5	6.9	247	4	US-09-902-540-13374	Sequence 12374, A
38	232	6.9	256	3	US-09-134-001C-4600	Sequence 4600, Ap
39	228.5	6.8	222	4	US-09-602-787A-442	Sequence 442, App
40	228	6.8	361	4	US-09-107-532A-6490	Sequence 6490, Ap
41	227.5	6.8	252	4	US-09-902-540-12415	Sequence 12415, A
42	227.5	6.8	273	3	US-09-134-001C-3641	Sequence 3641, Ap
43	227.5	6.8	294	4	US-09-583-110-3579	Sequence 3579, Ap
44	226.5	6.8	1302	1	US-08-232-537-2	Sequence 2, Appl1
45	225.5	6.7	393	4	US-09-248-796A-20209	Sequence 20209, A

#### ALIGNMENTS

RESULT 1									
US-09-245-808-1									
Sequence 1, Application US/09245808									
Patent No. 6313277									
GENERAL INFORMATION:									
APPLICANT: Doyle, L. Austin									
APPLICANT: Abruzzo, Lynne V.									
APPLICANT: Rose, Douglas D.									
TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which									
TITLE OF INVENTION: encodes it									
FILE REFERENCE: Rose Umb conversion									
CURRENT APPLICATION NUMBER: US/09/245,808									
CURRENT FILING DATE: 1999-02-05									
EARLIER APPLICATION NUMBER: 60/073763									
EARLIER FILING DATE: 1998-02-05									
NUMBER OF SEQ ID NOS: 7									
SOFTWARE: PatentIn Ver. 2.0									
SEQ ID NO 1									
LENGTH: 655									
TYPE: PRT									
ORGANISM: Human MCF-7/AdrVp cells									
US-09-245-808-1									
Query Match									
Best Local Similarity 100.0%; Score 3352; DB 3; Length 655;									
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MSSSVVEVIFPVSGQNTNGFPATASNDLKATFEGAVLSFPHNICRYVKLSGFLPKRPVE	60						
DB	1	MSSSVVEVIFPVSGQNTNGFPATASNDLKATFEGAVLSFPHNICRYVKLSGFLPKRPVE	60						
QY	61	KEILSNINGIKMPCGLNAILGPTGGGKSLDLVLAARKDPSGISGDLINGADRPANFKCN	120						
DB	61	KEILSNINGIKMPCGLNAILGPTGGGKSLDLVLAARKDPSGISGDLINGADRPANFKCN	120						
QY	121	SGYVAVDDVVMGTLTVRENLQPSAALRLATTMTNHEKNERINRVIOELGLDLYADSKVGT	180						
DB	121	SGYVAVDDVVMGTLTVRENLQPSAALRLATTMTNHEKNERINRVIOELGLDLYADSKVGT	180						
QY	181	QFIRGVSGGGERKRSIGMELTTDPSILFLDEPTTGLDSTANAVLLLRKMSKOGRTIIF	240						
DB	181	QFIRGVSGGGERKRSIGMELTTDPSILFLDEPTTGLDSTANAVLLLRKMSKOGRTIIF	240						
QY	241	SIHQPRYSIFKLPDLSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADFLDIING	300						
DB	241	SIHQPRYSIFKLPDLSLTLLASGRLMFHGPAQALGFESAGYHCEAYNNPADFLDIING	300						
QY	301	DSTAVALNREDDFKATETIEPSKODKPLIEKLAETIYNNSSFKETKALHQLHSGGEKKKK	360						
DB	301	DSTAVALNREDDFKATETIEPSKODKPLIEKLAETIYNNSSFKETKALHQLHSGGEKKKK	360						

```
QY 361 ITVFEKISTYTSFCHOLRWYSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND5
DB 361 ITVFEKISTYTSFCHOLRWYSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND5
QY 421 TGIQNRAGVLFELTTNOCSSSVSAVELFVVEKKLFHEHYSIGYRVSSTYFGLKLSDLPL
DB 421 TGIQNRAGVLFELTTNOCSSSVSAVELFVVEKKLFHEHYSIGYRVSSTYFGLKLSDLPL
QY 481 MTMLPSIIIFTCIYFVGLKPKADAFVVMFTLMVAVYASASMAAIAAGQSVSVATLL
DB 481 MTMLPSIIIFTCIYFVGLKPKADAFVVMFTLMVAVYASASMAAIAAGQSVSVATLL
QY 541 MTICFVPMIFSGLLVNLTTIASWLSWLOFSPRIGFTALQNEHFLQGFCEGLNATGN
DB 541 MTICFVPMIFSGLLVNLTTIASWLSWLOFSPRIGFTALQNEHFLQGFCEGLNATGN
QY 601 NPCNYATCTGEEYLVKQIDLSFPGMLKKNHVALACMIVFLTAIYKLLFLKKYS
DB 601 NPCNYATCTGEEYLVKQIDLSFPGMLKKNHVALACMIVFLTAIYKLLFLKKYS
```

## RESULT 2

```
US-09-167-594-1
; Sequence 1, Application US/09767594
; Patent No. 6521635
; GENERAL INFORMATION:
; APPLICANT: Bates, Susan
; APPLICANT: Robey, Robert
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives
; FILE REFERENCE: 015280-402100US
; CURRENT APPLICATION NUMBER: US/09/767, 594
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 60/177, 410
; PRIOR FILING DATE: 2000-01-20
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human mitoxanthrone resistance (MXR)/BRCP/ABCP
; OTHER INFORMATION: protein
US-09-167-594-1
```

Query Match 99.4%; Score 3331; DB 4; Length 655;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MSSSNVEFIPVSOQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSKGFPCRKPYE
DB 1 MSSSNVEFIPVSOQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSKGFPCRKPYE
QY 61 KEILSNINGIMKGLNALIGPTGGKSSLLDVLAAKDPGSLGSDVILNAGAPPANFKN
DB 61 KEILSNINGIMKGLNALIGPTGGKSSLLDVLAAKDPGSLGSDVILNAGAPPANFKN
QY 121 SGVYVDDVVMGTLVRENLOFSALRLATMTNHEKNERINRYIOELGLDKVADSKVGT
DB 121 SGVYVDDVVMGTLVRENLOFSALRLATMTNHEKNERINRYIOELGLDKVADSKVGT
QY 181 QIFRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANAVLLKMSKQGRITIF
DB 181 QIFRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANAVLLKMSKQGRITIF
QY 241 SIHQPRYSIFKLPDSLTLLASGRIMFHPAQOALGYFESAGYHCEAVNNPADPFLDING
DB 241 SIHQPRYSIFKLPDSLTLLASGRIMFHPAQOALGYFESAGYHCEAVNNPADPFLDING
```

```
QY 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNNSFYKETAEHLHOLSGEKKKK
DB 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNNSFYKETAEHLHOLSGEKKKK
QY 361 ITVFEKISTYTSFCHOLRWYSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND5
DB 361 ITVFEKISTYTSFCHOLRWYSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND5
QY 421 TGIQNRAGVLFELTTNOCSSSVSAVELFVVEKKLFHEHYSIGYRVSSTYFGLKLSDLPL
DB 421 TGIQNRAGVLFELTTNOCSSSVSAVELFVVEKKLFHEHYSIGYRVSSTYFGLKLSDLPL
QY 481 MTMLPSIIIFTCIYFVGLKPKADAFVVMFTLMVAVYASASMAAIAAGQSVSVATLL
DB 481 MTMLPSIIIFTCIYFVGLKPKADAFVVMFTLMVAVYASASMAAIAAGQSVSVATLL
QY 541 MTICFVPMIFSGLLVNLTTIASWLSWLOFSPRIGFTALQNEHFLQGFCEGLNATGN
DB 541 MTICFVPMIFSGLLVNLTTIASWLSWLOFSPRIGFTALQNEHFLQGFCEGLNATGN
QY 601 NPCNYATCTGEEYLVKQIDLSFPGMLKKNHVALACMIVFLTAIYKLLFLKKYS
DB 601 NPCNYATCTGEEYLVKQIDLSFPGMLKKNHVALACMIVFLTAIYKLLFLKKYS
```

## RESULT 3

```
US-09-538-092-72
; Sequence 72, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538, 092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127, 352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178, 965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapacSeqFormer Version 0.9
; SEQ ID NO 72
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YCR011C
US-09-538-092-72
```

Query Match 24.9%; Score 835.5; DB 4; Length 1049;  
Best Local Similarity 30.5%; Pred. No. 3e-77;  
Matches 222; Conservative 134; Mismatches 257; Indels 115; Gaps 18;

```
QY 1 MSSSNVEFIPVSOQNTNGFPATASNDLKAFTGAVLSFNHCYRVKLSKGFPCRKPYE
DB 355 LGSXSKSPILRP-DEDAVNNFLQNEBDTL-----ALTSEENITVYSPSINS-----DGYE
QY 61 KEILSNINGIMKGLNALIGPTGGKSSLLDVLAAKDPGSLGSDVILNAGAPPANFKN
DB 61 KEILSNINGIMKGLNALIGPTGGKSSLLDVLAAKDPGSLGSDVILNAGAPPANFKN
QY 119 CNGSYVDDVVMGTLVRENLOFSALRLATMTNHEKNERINRYIOELGLDKVADSKV
DB 463 KIIGFVDDDFLPTLVETVINSALRLPKALSEAKKARYKVLBEIRIIDIIDRIT
QY 179 GTOFIRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANAVLLKMSKQGRITIF
DB 523 GNEFDRGISGGERKRTSIGMELITDPSILFDEPTTGDDSTANAVLLKMSKQGRITIF
QY 238 IIFSIHQPRYSIFKLPDSLTLLASGRIMFHPAQOALGYFESAGYHCEAVNNPADPFLDI
```

Db 583 LVLSTHQPNSNFYLPDKLVLLSKGEMVYSGNAKKVSEFLRNBEGYICPDNVIADVLIDI 642  
QY 298 -----INDSTAV 305  
Db 643 TFEAPQCKRRIRINISDLAECTDNDNTIHQTTFTSSDGTTOREMAHLAHLDEIRS 702  
QY 306 ALNREEDFKATE-----IIEPSKODKPLIEKLAIEIYVNSGFYKETAELHQ-LSGEEKKK 360  
Db 703 LLRDEEDVBGTGRGATETIDNTKLLHDK-----XDSVYVYELSGEIEVUSSEGEEN 758  
QY 361 IT-VFKEISYTTSCFCHQRMWSKSGFKNLLGNPOASIAQIIVTVVLGLVIGAIYFGLKQ 418  
Db 759 VLANGDLPTGQAGAFLOQLSILNSRSFKWMYRNPKLLNGVLLTLLSFLGTLYVNVN 818  
QY 419 DSTGIONAGVLPFLTTNOCSSVASAVELFVVEKLFTHIEYISGYRVSYFLGKLLSL 478  
Db 819 DISGQNRKGLFELTYEGFVTFTGLSSPALERIIPIKERNNYSPPLAAYISKIMSEV 878  
QY 479 LPMTMLPSIIFTCIYVFMGLKPKADAFVMMFTLMMVAVSASSMALAIAGQSVSVAT 538  
Db 879 VFLRVVPLLSLIYPMGLMKDPAFKCIGILIFMLGISLELITIGIFEDLANSI 938  
QY 539 LMTICFVFMIFSGLLV---NLTTIASWLMLOYFSIPRYGFTALQNEF----- 586  
Db 939 ILSVLVLGSLFLSGFLPINTKNITNVA--FKYLNQPSVYVAVESLLINEVKTMLKEX 996  
QY 587 LGQNP-CRGLNATGNPNPCVATCTGEEYLVKQI--DLSPLMKHVALACMIYFLTI 643  
Db 997 YGLNTEVRP-----ATILSTFGFVQNLVVDIK-----ILALENVFLIM 1036  
QY 644 AYKLLFL 651  
Db 1037 GYLLKMI 1044

RESULT 4  
US-09-538-092-1125  
; Sequence 1125, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Glot, Loic  
; APPLICANT: Manafield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatsqFormatter Version 0.9  
; SEQ ID NO 1125  
; LENGTH: 674  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P45844  
US-09-538-092-1125

Query Match 21.1%; Score 706.5; DB 4; Length 674;  
Best Local Similarity 28.4%; Pred. No. 3,9e-64;  
Matches 194; Conservative 155; Mismatches 251; Indels 83; Gaps 23;  
QY 3 SGNVVFIPVSGNNGPATAISNDL---KATF---EAV-LSFHNICRYKLSGRLP 54  
Db 34 SSNMBA---TETDLNGLHKVKNNTTEAQRFSPLPRRAAVIETPRDSYSVPEGPW-- 88  
QY 55 CRKPKETILSINGIMKPG-LNAILGPTGCGKSLDLVLAARKDPSGLSGDVLINGAPR 113

Db 89 -RKKGYKTLKIGSKFNSGELVIAIMPSGACKSTLMNIIAGYRE-TGNKGAVLINGLPR 146  
QY 114 PAN-FKNSGYVODDVMTGTLVRENLOFSALRLATTTMHEKNERINRYOELGDK 172  
Db 147 DLRCKRKSICYMODMLPHLTVQEMVNSHLQOE--KQGRREMKETILTALGLS 204  
QY 173 VADSKVGTQPIRGVSGRKRKTSIGMELITDPSILFDEPTTGDSSTANAVALLLKMS 232  
Db 205 CANTRTGS-----LSGQGRKRLAIALBELVNNPVPWFPEPTGSLDASCFQVSLMKGLA 259  
QY 223 KQGRITIFSIIHQPRYSIFKLPSLTLASGRIMFGPAOALGYFESAGYHCEATYNNPD 292  
Db 260 QGSRSIITIHQPSAKLPFLPDQLVLSQGCQCYRGKVCNLPYRDGLNCPYHNPAD 319  
QY 223 PFLDIINDSTVALNREEDFKATEIIEPSKODKPLIEKLAIEIYVNSGFYKETAEL-KAEL- 349  
Db 320 FMEVAVSG-----EYGDNSRLVRAVRGECDSHKKRLGDAEYN 360  
QY 350 ---HQLSGEKK-KKITVFEKISYTTSCFCH-----QLRWKRSFKNLLGNPOASI 396  
Db 361 PELMRPBEVYQTRKLGKRDSSMEGCHSFASCLQPCILFKRTFLSIMRDSVLT 420  
QY 397 AOIIVTVGLVIGAIYGLKNDSTGIONRAGVLPFLTTNOCSSVASAVEL-FVVEKLP 455  
Db 421 LRITSHIGILIGLILYIGIGNEAKKVLNSNGFLFPMFALMPTVLTPLMEGVF 480  
QY 456 IHEIYSGYRVSSYFLGKLLSDLLPMTMLPSIIFTCIYVFMGLKPKADAFVMMFTL-M 514  
Db 481 LREHLNMYSLKAVYLAKTMD-VPPQIMFPAYCSIVYMMTS--OPSDAVRVLPAALDT 538  
QY 515 MVAASASSMALAIAAGQSVSVATLMTICFVFMIFSGLLVNLTTIASWLMLOYFSIP 574  
Db 539 MSLVAQSLGILGASASTSQVATFVGPVTALPVLLFSGFPSPTIPLYLQMSYISV 598  
QY 575 RYGF-----ALQNEFLGQPCPGANGNPNPCVATCTGEEYLVKQIDLSPLGIM 627  
Db 599 RYGFEGVILSYGLRED-----LHCDIDETCHQK--SEAILRE-LDVENAKLY 645  
QY 628 KHVVALACMIYFLTIAYLKLF 650  
Db 646 LDFVLGIFFLSLRLIAYFLRY 668

RESULT 5  
US-09-989-981A-2  
; Sequence 2, Application US/09989981A  
; Patent No. 6821750  
; GENERAL INFORMATION:  
; APPLICANT: Hobbs, Helen H.  
; APPLICANT: Shan, Bei  
; APPLICANT: Barnes, Robert  
; APPLICANT: Tian, Hui  
; APPLICANT: Tularik Inc.  
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use  
; FILE REFERENCE: 018781-007320US  
; CURRENT FILING DATE: 2002-07-23  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 652  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: Mus musculus  
; OTHER INFORMATION: mouse ABCG5 (mABCG5)  
US-09-989-981A-2  
Query Match 20.7%; Score 693.5; DB 4; Length 652;



Best Local Similarity 29.0%; Pred. No. 8.4e-63;  
Matches 181; Conservative 142; Mismatches 246; Indels 55; Gaps 16;

```

Oy      12  JSQKNTNGEPAYASNDILKAFTEGAVUSFPHNI CYRVKLKSG-----FLPCRPVKEKILSNI 67
Db      25  LEQSGSVTGEAHS-----LGVLHVSISVSNRRQPMWNITSCQKMRQILKDV 73
Oy      68  NGIMKPG-LNALIGPTGGKSSLLDVLAARKDPSG-LSGDVLINGAP-REPAHEKNSGYV 124
Db      74  SLVIESGQIMCIISSSGSGKTTLLDAISGLRARTGLEGEVFPVNGCELRDRDQPCPSYV 133
Oy      125  VODDVMGTLTYRENLFQSAALRLATTTMNHENESINRYIOELGLDKVADSKVGOQFIR 184
Db      134  LQSDVFLSSLTYRETLRYTAMALACRSSADF-YNKKVEAAMTELISLSHADOMIGSNFG 192
Oy      185  GVSQGERKRTSIGMELITPSSILFLDEPTGLDOSTANAVLILKMKSKGRTIFSIHQ 244
Db      193  GISSERRRVSIAAQLODPKXMMDEPTGLDQMTANQVLLAEIARDRIVITIHQ 252
Oy      245  PRYSIFKLPDSLTLASGRLMFHFGPAQOALGYFESAGYHCEAVNPNADFLDIINGDSTA 304
Db      253  PRSELFQHDKALAILTYGELVFCGPTEBEMIGFENNNGYPCPEHSNPFDPMDLTSDVTO- 311
Oy      305  VALNNE-BDEKATEIIEPSKQKPLBKLAIEIVVNSFYKETAELHQLSGGKKKITY 363
Db      312  -SRREIEIYKAVOMECAFKESDIYHKI-----LENIIRARLKT 351
Oy      364  FKEISYTT---SFCQHLRWMSKRSPKNLIGNPOASIAQIIYTVVLGVLGAIYFGLKND 419
Db      352  LPMVFPKTDPPGMFEGKGLVLLRVTNRINMRNQAYIMRLVQNLINGLFLIFLLRQNN 411
Oy      420  ST--GIQNGAVLFLITNQCFSS-VSAYELVEVEKKLFHEYISGYRVSSYFLGLLS 476
Db      412  TLKGAVDORVGLLYQLVGATPYTGMINAVLFLPMLRAVSDOEDGILYHQMILLAYVL- 470
Oy      477  DLEPMTMLPSIIFETCIVYFMLGIKPRADAFFVMMFLM---WVAYSASSMALIAAGQSV 533
Db      471  HVLPESVIATVLFSSVCTMLGLYFPAVAFGYFSAAALBAPHLIGEFITLVLLGIVQNPNI 530
Oy      534  VSVATILMTICFVPMNIFSGLLVNLTTIASWLSMLOFYSI-PRYGFPTALQHNEFLQONF-C 592
Db      531  VNSIVALLSIS--GLLIGSGFIRNIQEMDPIKILIGYFTQKVCCEBILVNEFYGLNFTC 588
Oy      593  PGLNATGNPNPCVATCTGEEYLVK 616
Db      589  GGSNTSNLHMPCCALITQGVQFLEK 612

```

```

/ RESULT 6
/ US-09-989-981A-6
/ Sequence 6, Application US/0989981A
/ Patent No. 6821750
/ GENERAL INFORMATION:
/ APPLICANT: Hobbs, Helen H.
/ APPLICANT: Shan, Bei
/ APPLICANT: Barnes, Robert
/ APPLICANT: Tian, Hui
/ APPLICANT: Tularik Inc.
/ APPLICANT: Board of Regents, The University of Texas System
/ TITLE OF INVENTION: ABC05 and ABC08: Compositions and Methods of Use
/ FILE REFERENCE: 018761-007320US
/ CURRENT APPLICATION NUMBER: US/09/989,981A
/ CURRENT FILING DATE: 2002-07-23
/ PRIOR APPLICATION NUMBER: US 60/252,235
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 60/253,645
/ PRIOR FILING DATE: 2000-11-28
/ NUMBER OF SEQ ID NOS: 13
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 6
/ LENGTH: 651
/ TYPE: PRT
/ ORGANISM: Homo sapiens

```

```

; FEATURE:
; OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981A-6

```

```

Query Match 20.4%; Score 682.5; DB 4; Length 651;
Best Local Similarity 29.2%; Pred. No. 1.2e-61;
Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;

QY 13 SGGNTNGFPATASNDLKAFTGGAVALSPHNICRYVKLSGFLPCRKPEVEKILSINGIMK 72
D 21 SSSLEGAAPATP---EPHSIGILHASVSHVRPMMWDITSCQOMTRQILKDVSLYVE 77
QY 73 PG-LNALLPFGGKSSLLDYLARKPSG-LSGDVYLNG-APPANPKCKSVYVDV 129
D 78 SGOIMCILGSSGSKTLLDAMSRLRAGFLGSEVYVNGRALRREOFQDFCSYVLSDDT 137
QY 130 VMGILTYRENIQFSAALRLATTMTNHEKNERINEIVIOELGIDKADSKSVGTQIRGVS 189
D 138 LLSLTYRETHYRLALLAIRGNPG-SFOKKEVAVMELSLSHADRLIGNYSLGISTG 196
QY 190 ERKETSIGMELITDPSILFLDEPTTGIDSTANAVALLLKMSKQRTTIFSIHQPRYSI 249
D 197 ERRRVSAIAQLQDPKMWLEDEPTTGIDCMTANOIVLVELARNRHIVLTIHQPRESL 256
QY 250 FKLEDSILTLASGELMFHQAQELAGFEASGHCBAVNNPADPFLIDINGDSTAVALNR 309
D 257 FQLPDKTALISFGLIFCGTPAEMLDFFPDGFCPEHSPNFDYMDLTSVDTO---SK 312
QY 310 EEDPKATEIIEPSKQDPLIEKLAIEIVYNSGFYEKTRAEALHQLSGEKKKKITVEKLSY 369
D 313 ERE-----IETSKR-----VQMISAVKKAICHKT-----LNINEMKMLKTLPMWPF 356
QY 370 TT---SFCHOLRWVSKRSEFNKLGNPOASTAQIITIVVLGVLGAIYFGKANDST---- 421
D 357 KTKDSPGVFSKLGVLRLRVTRNLRNKLAVITRLLOLIMGLFL--LEFVLIRVSNVLKG 414
QY 422 GIORAGVLF-FLTNQCFSSVASVELFVVEKKLFIHEYISGYRVSSYFGLKLSLDLP 480
D 415 AIDQRVGLVQFVATRYTGTMLNANVLPFVLRAVSDDESQDGLYQKQOMLALYAL-HLP 473
QY 481 MTMLPSIIFTCIYFMLGLKPKADAFVMMFTLM--MVAYSASSMALAIAGGSVVSVA 537
D 474 FSVVATMIFSSVCWTGTLGHEVARFGFSAALLAPHLIGEFLLTVLLGIIQONRIYVNSV 533
QY 538 TLLMTICFVPMIRSGILVNLTTIASMLSMVQVESIDRYGTALQHNFEFLQNFPCG--- 594
D 534 VALISLIGV--LVGSGFLRNIGQPIIPKTIISYTFPQRCSEILVNEFYGLNFTCGSSN 591
QY 595 LMATGNPNPCNYA-----TCTG 610
D 592 VSVTTNPMCAFTQGIQFIKTCPG 615

RESULT 7
US-09-989-981A-4
Sequence 4, Application US/0989981A
Patent No. 6821750
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bei
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularek Inc.
APPLICANT: Board of Regents, The University of Texas System
TITLE OF INVENTION: ABCGS and ABCGS: Compositions and Methods of Use
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989, 981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252, 235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253, 645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13

```

SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 672  
TYPE: PRT  
ORGANISM: Mus musculus  
FEATURE:  
OTHER INFORMATION: mouse ABCG8 (mABCG8)  
US-09-989-981A-4

Query Match 19.6%; Score 657.5; DB 4; Length 672;  
Best Local Similarity 27.2%; Pred. No. 5e-57;  
Matches 185; Conservative 141; Mismatches 270; Indels 85; Gaps 21;

```
20 PPTASNDLKAFET---EGAVLSFNHCYRVKLSGKGF-----LPCKRKEKEI--- 63
28 PPSSESDNSL-YFTYSGQSWTLEVRDLTYQVDIASQVWPEQLAQFIRPKSHSSQDSCELT 86
64 -LSNNGIMKPG-LNAIIGPTGGKSSLLDVLAARKDPSSL-SGDVLINGAPRPANF--K 118
87 GIRNLSPKVRSGOMLALIGSSGGRASLLDVITGRHGKMKSGQIMWINGQSTPQLVYRK 146
119 CMSGVYVDDVVMGTLTVRENIQFSAALRLATMTNHEKNERINRYIOELGLDKVADSKY 178
147 C-VAHVROHDLLPMLTVARETLAFIAQMKLPRTFSQAQRDKVEDVIAELRLRQCANTRV 205
179 GTPFRGVSGGRRKRTSIGMELITDPSILFDEPTTGLDSTANAVILLKMKMSKQRTI 238
206 GNTTVRGVSGGRRKRTSIGMELITDPSILFDEPTTGLDSTANAVILLKMKMSKQRTI 265
239 IFSIHQPRYSIFKLPDSLTLLASGRLMFGRPAQOALGPESAGYHCEAYNPNADPFLDI 298
266 LISLHQPRSDIFRLFDVLVMTSGPTIYGAAGQWQVFTSIGHPCPRYSNPADFYVDLT 325
299 NQDSTVALNREDEKATEIIEPSKODKFLIEKLAETVYN-----SSFYKETKALHQIS 353
326 SIDRS-----KEREVATVEK-----AQSLALFLEKVOGQPDFL--WKAERKEIN 369
354 GGEKKKKITVPKEISYTTSS-----FCHOLRWVSKRSPKULLGNPOASIAQIIVTVVLGV 408
370 TSTHTVSLTLTODTDCGTAVELPGMIEBPSTLIRQISNDPRDLPTLLHSGSEACLMSTLT 429
409 IGAIVFGKNDSTGIONRAGVLFPL-----TTNOCFSSVSAVELFVVEKKLFTIH 457
430 IGFLLYGGAKQSLSDMTDRLALFMIGALIPNVILDVYSKCHS-----ERSMLYX 479
458 EYISGYRVSSYFLKLSLDLIPMTLPSIIFTCIVYFMLGLKPKADAFVVMFTLMNYA 517
480 ELEDGLYTAGPYFFPAKILGE-LPEHCAYVILIYAMPYIYMLTNLRPVBELFLHFLVWLIV 538
518 YSASSMALALIAAGOSVSVATLMTICFPMIMFSGLVNLTIASWL--SMLOYFESTPR 575
539 FCCRMALASAMLPTEPHMSFFCNALYNSFYLTGFMNLDNL--WIVPAVISKLSFLR 596
576 YGFALQHNPELFGONFCGLNATGNPNCAVYA--TCTGEETVLVQGIIDLSPMGLMKHVALA 634
597 WQFSLMQQIFGHILYTTQIG-----NFTTSIIIGDTMI--SANDLNSHPLIYALYIVI 647
635 CMIVIFLTIAYLKLFLKKYS 655
648 GISYGFLLPYLVLKLIKOKS 668
```

RESULT 8  
US-09-989-981A-8  
Sequence 8, Application US/0998981A  
Patent No. 6821750  
GENERAL INFORMATION:  
APPLICANT: Hobbs, Helen H.  
APPLICANT: Shan, Bei  
APPLICANT: Barnes, Robert  
APPLICANT: Tian, Hui  
APPLICANT: Tularik Inc.  
APPLICANT: Board of Regents, The University of Texas System

TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use  
FILE REFERENCE: 018781-007320US  
CURRENT APPLICATION NUMBER: US/09/989,981A  
CURRENT FILING DATE: 2002-07-23  
PRIOR APPLICATION NUMBER: US 60/252,235  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/253,645  
PRIOR FILING DATE: 2000-11-26  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8

Query Match 19.1%; Score 640.5; DB 4; Length 673;  
Best Local Similarity 27.2%; Pred. No. 3e-57;  
Matches 187; Conservative 139; Mismatches 273; Indels 89; Gaps 21;

```
16 NTNG-----PPTASNDLKAFTEGA--VLSFNHCYRVKLSGKGF-----LP---- 54
19 DTSGIQDRLFFSSESDNSLYFTYSGQPNLTLEVDLNYQVDIASQVWPEQLAQFKMPWTSR 78
55 -CRKVEKEILSNNGIMKPG-LNAIIGPTGGKSSLLDVLAARKDPSSL-SGDVLINGA 111
79 SCNSCEIGET-QNLSFKVRSGOMLALIGSSGGRASLLDVITGRHGKMKSGQIMWINGQ 137
112 PRPNANF--KMSGVYVDDVVMGTLTVRENIQFSAALRLATMTNHEKNERINRYIOELG 169
138 PSSPOLVYKC-VAHVROHDLLPMLTVARETLAFIAQMKLPRTFSQAQRDKVEDVIAELR 196
170 LDKVADSKVQTPFRGVSGGRRKRTSIGMELITDPSILFDEPTTGLDSTANAVILLIK 229
197 LRQCADTVGMVYVGLSGGRRKRTSIGMELITDPSILFDEPTTGLDSTANAVILLIK 256
220 RMSKQRTIIFSIHQPRYSIFKLPDSLTLLASGRLMFGRPAQOALGPESAGYHCEAYNPN 289
257 RLAKGNRLVLSLHQPRSDIFRLFDVLVMTSGPTIYGAAGQWQVFTALIGPCPRYSN 316
290 PADFLDIINGDSTVALNREDEKATEIIEPSKODKFLIEKLAETVYNSSFYK-ETKAE 348
317 PADFLVDLTSIDR-----SREDELATRE--KAQSLALFLEKVRPL--DDEFLMKATK-- 366
349 LHOISGGEKKKKITVPKEISYTTSPC-----HQLRWVSKRSPKULLGNPOASIA 397
367 -----DLDEDTCVESVTPPLDNTCLPSPTMPCGAVOQFTLLIRQISNDPRDLPTLLI 419
398 QIIVTVVLGVIGALIFGLKNDSTGIONRAGVLFPL-----TTNOCFSSVSAVE 446
420 HGAECALMSMTIGFLYFGHSGIQLSFMDTALLFMIGALIPNVILDVYSKYS----- 473
447 LFVVEKKLFIHEYISGYRVSSYFLKLSLDLIPMTLPSIIFTCIVYFMLGLKPKADAF 506
474 ----ERANLYLEBEGLYTTGPIYFFAKILGE-LPEHCAYIILIYGMPTYMLANLRGLQPF 528
507 FVMFTLMVAVYSSSMALALIAAGOSVSVATLMTICFPMIMFSGLVNLTIASWLS 566
529 LLHFLVLVLFCCRIMALAALALLPTHMASFFSNALYNSFYLAGGFMINLSLMTVPA 588
567 WLOYFSIRYGFALQHNPELFGON--CGLNATGNPNCAVATCTGEETVLVQGIIDLSPMG 625
589 WISKVSFLRWCPEGLMKIQFSRRYKMLGULT-----IAVSGDKIL--SAMEDSYR 639
626 LMKHVALACIIVIFLTIAYLKLFLKK 653
640 LYALYIYIGSGGFMVLYVLSLRIRIQ 667
```

RESULT 9  
US-09-614-912-138



```

Db      926 VYISQGLGRNSQKMEYFEALPGVPKIKDKY-----NPATMWLEV---SSVAAEVRLNM 976
Qy      312 DPKAIEIIEPSKQDKPLIEKLAIEIYNSFFYKETAELHQLSGGKKKIITYFKELSYTT 371
Db      977 DF-----ABEYKTSIDLKQNKVLVNLQSPBEGTSDLHF-PTKXG 1016
Qy      372 SFCHOLRWYSKRSFKNLGNPQASIAQIIVTVVLGVIGAIYF--GLK-NDSTGIONRAG 428
Db      1017 STIGCFRACLKQKWLTYRSFDPYNLVRSFTLFTALLIGTITWKIGTKGNANSLRWYIG 1076
Qy      429 VLF-----FLTINQCFSSVAVELFVVEKKLFIEHYSIGYRVSSYFLGLSDLLPMTWL 484
Db      1077 AMYAVMFIGINNC---ATVQPIVSIERTVFRERAGMYSAMPYALQVME-IPYVVF 1132
Qy      485 PSIIITCIYFEMGLKPKKADA---FFVMFTLMVAVSASSMALAIAGQSVSVATLL 540
Db      1133 QTAAYVTLIVYAMMSQMTAKKFFWFFVSYFSPFLYFTY-YGMWTVASISPH--EVAALF 1188
Qy      541 MTCFVFMIFSGLLVNLTTISWLSWLOYSIPR---YGFALQHNE 585
Db      1189 AAAYSLFNLFSRFFIIPRPIKMWIM-YWLCPLAMTVYGLITVQYGD 1236

```

# RESULT 11

```

US-09-248-796A-20212
; Sequence 20212, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20212
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (740)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-248-796A-20212

```

```

Query Match      15.4%; Score 516; DB 4; Length 740;
Best Local Similarity 26.7%; Pred. No. 3.3e-44;
Matches 169; Conservative 106; Mismatches 245; Indels 112; Gaps 21;

Qy      16 NNGEPATASNDLAKFTSGAVLSPHNICRVLYKSGFLPCRKVEK---ELISNINGIMK 72
Db      135 NDSDDQPTVTNALMFTTEAL-----NKLK-----KPDSSKYFDILKSDMALIR 178
Qy      73 PG-LNALIGPTGGKSSLDVLAAR-----KDP-----SGLS-----GDVLINGA 111
Db      179 PGEILVIVGRPGAGSTILKTIAMVTYGFHIGKEQITDGLSPHDIERNHGVDI----- 234
Qy      112 PRPAFKNSGVVDDVVMGTLTYRENLOFSALRLATTMTNHEKNERIN----- 163
Db      235 -----YSAETDVHFPRLSVGDTLEFAARL-----TPQNRGEGIDRETYAKMA 278
Qy      164 ---VIGELGLDKVADKVGQFIRGVSGGERKRTSIGMELITDPSILFLDEPTGIDGSTA 221
Db      279 SVYMAVTYGSHTRNNTNNGDNDFRGVSGGERKRVSIABSLGANTQCWMDNATRLCLDSATA 338
Qy      222 NAVLLLLKRMKSGRTIIFSIHOPRYSIFKFLFDSLTLASGRLLMFHGPACALGYFESA 280
Db      339 LEFIALKTSATILDTTLPILAIYQCSQAAYELFDVNVVLYVEGYQIFPGKASAKRYFENM 398

```

```

Qy      281 GYHCEAVNNPADPFLDINGDSTAVALNREEDFKATEIIEPSKQD--PLIEKLAIEYVN 338
Db      399 GWKCFQROQTADFLSLTN-----PAEREP-----LPGVEDKVPRTAQGEFFMGN 444
Qy      339 SPSFYKETAELHQL-----SGGKKKIITYFKELIS-----YTSFCHOLRWYSKRS 384
Db      445 PEVEALTEIIDEYCECERSNTGETYRESHVAQSSNTRPSSPTVSPFMQRYVIAN 504
Qy      385 FKNLGNPQASIAQIIVTVVLGVIGAIYFGKANDSTGIONAGVLFLLTNQCFSS-VS 443
Db      505 FLRMGDBRSIHIFSYFQGLWMLIISVFYNLSQTTGSFYRGAMIFAVLENAPSSSLE 564
Qy      444 AVELF-----VVEKKLFIEHYSIGYRVSSYFLGLSDLLPMTWLPSIIFTCIYFEMGL 499
Db      565 IWSLEAPRIYEK---HKYA-FIRPADALASIISE-LPVKXLSNBSFNVFIFMVNF 618
Qy      500 KPKADAFVMMFTLMVAVSASSMALAIAGQSVSVATLLMTICFVFMIFSGLLVNL 559
Db      619 RNPGRFFPYMLMCWCFWVSHLFRSIGAVSTISGAMTPATVLLANVIYTGFIPTP 678
Qy      560 TIASWLSWLOYSIPRYGFTALQHNEFLQNF 591
Db      679 SWLGSRWIMYINIPGVYFESILMVEFHGREF 710

```

# RESULT 12

```

US-09-248-796A-20772
; Sequence 20772, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20772
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20772

```

```

Query Match      15.3%; Score 514.5; DB 4; Length 518;
Best Local Similarity 29.4%; Pred. No. 2.6e-44;
Matches 161; Conservative 112; Mismatches 217; Indels 57; Gaps 22;

Qy      121 SGVYVQ-DVVMGTLTYRENLOFSALRLATTMTNHEKNER-INRVIQELGLDKVADSKV 178
Db      15 SAYGQHSPLIEQLTVARETLTYQAKLRL--PLDQHKFIPITINKLIGTGLVDCADTLI 72
Qy      179 GTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTGIDGSTANAVLLLLKRMKSGQR-T 237
Db      73 GSEYVKGISGGERKRVSIQILSLKPKVLFDEPTSGIDSSYAEVITLLGLAKENMTT 132
Qy      238 IIFSIIHOPRYSIFKFLFDSLTLA--SGRLMFGPQAPBALGYFESAGYHCEAVNNPADPFLD 296
Db      133 IILTHQPSQULFYFGSLLLGRGKVIYDTSIGIYELSDGIYNNPEGNINDYILD 192
Qy      297 IINGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYNSFFYKETAELHQLSGGE 356
Db      193 LISRMENDKQGLER--RVAEIL--SYQANSIKKLGCS---TATPLQEI-IDLPQY----- 240
Qy      357 KKKKITVYKEISYTTSPCHOLRWYSKRSFKNLGNPQASIAQIIVTVVLGVIGAIYFGL 416
Db      241 YYORLPIF--ITFPIFRQL--LTSYRAKQVNVINRACQ-----TIFLAIYHTLYFPTPL 290
Qy      417 KNDSTGIONRAGVLFLLTNQCFSSVAVELFVVEKKLFIEHYSIGYRVSSYFLGLKLS 476

```

Db 291 RNTQEGISNRLGVOEVLNLYFAGLININILTYPERNLFOEYRDGIVGTVERGLSYLIN 350  
QY 477 DLLPMMLPSIIITCIYVFMGLKPKRADAFFVMMFTLMMVAYGASSAALAIAGQSVSV 536  
Db 351 E-LPTETVTPGFPAALVFAAGLPRTPOMFPAFGTGVINCESIGITVNSITFTMGV 409  
QY 537 ATLMTICFVPMFMMIFSGILVNLFTTASWLSMLQYFSIPRYGFTALQHNFTLQNFPCGLN 596  
Db 410 ATNVLS-TPVSLAIFMCGTMSL-HMPGFKGINPISPMKAVAI CANLGFKNQSF----- 462  
QY 597 ATGNPNPN--YATC---TGEE---YLVKQIDLSFPMGLMKHVALACMIVIFLTAYLK 647  
Db 463 -----KCNSEAAOCLLTGDEVLVSYNMKNLGPVVGGL-----IGC-LVIYRVAAILS 510  
QY 648 LFLFKKY 654  
Db 511 IYRVK 517

## RESULT 13

US-09-538-092-413  
; Sequence 413, Application US/09538092  
; Patent No. 675314  
; GENERAL INFORMATION:  
; APPLICANT: Glot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538, 092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurataseqFormatter Version 0.9  
; SEQ ID NO 413  
; LENGTH: 1411  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number Y11013C  
US-09-538-092-413

Query Match 14.6%; Score 488.5; DB 4; Length 1411;  
Best Local Similarity 25.1%; Pred. No. 7.2e-41;  
Matches 170; Conservative 137; Mismatches 272; Indels 99; Gaps 24;  
QY 32 TEGAVISFNHCYRVKLSGFLPCRKVEKEILSNINGIMKPGILNALIGPTGGKSLD 91  
Db 746 TQGHVISMKNINNTTIG-----DKKLINDASGYISSGTYALMGSSGAKTTLN 793  
QY 92 VLAARDPGLSGDVLINGAP--RPAFKNSGVVDDVVMGTLTYREMLQFSALRLA 149  
Db 794 VLSQRFESGVVTEGLIDGQPLTNIDAFRSIGFVQGDVHLELTYRESLEISCVLR-- 851  
QY 150 TWTNHEKERNIRVIOELGLDVKVADSKVGTQPIRGVSGGERKRTSIGMELTDPG-ILF 208  
Db 852 -----GQDPRDYLGVNSILRLP-----SEKLVDLSPTQRKLSTIGVELVTKPSILLF 900  
QY 209 LDEPTGLDSTANAVALLKMSKOGRTIIFSIHQRYISIFLFDSTLL-ASGRLMFH 267  
Db 901 LDEPTGLDDEALVTVQFLKLSMOGALICTIHQSKSVISYFDNIYLLKRGGEVVF 960  
QY 268 GPQOALGYESAGYHC---EAYNPPADFLDIINGDSTAVALNREDFKATEIIP--- 321  
Db 961 GSPNACDIFVAIDRLITPREMDNPADVIDVVGSSSTNIPMDAEKPTSSKIDEPVSY 1020  
QY 322 SKQDKPLIEKLAIEYVNSFFYKETKALHQLSGGKKKKITVFKEISYTTTSFCHQLRWVS 381

Db 1021 HKQSDSI--NMALWQSSPEKRVVADDLLLEBEARKSGVDTTSVWSPSPSEIOIKLIT 1078  
QY 382 KRSFKNLGNPQASIOIITVTVGLVIGAIYGLKNDSTNGIONRAGVLEFLTNOQFSS 441  
Db 1079 KROYICTKRDMTYFAKYNALNAGALFIFGSEFRTHKNINGLODAFLCMMML-----CVSS 1135  
QY 442 --VSARELVPEVKEKLFIEHYISGYRVSSYFLCKLSDL---LPMTMLPSIIFTCIYVF 495  
Db 1136 PLINQODKALQSK---EVIYAREASNTFWHTVLLIAQIVELPLAISSTLIFLCYF 1192  
QY 496 MGLKPKRA--DAFVMMFTLMMVAYGASSAALAIAGQSVSVATLMTICFVPMFMMIFSG 553  
Db 1193 CCGFETTSARVAGVFNLYNLIFSMTYLSFGIMLLYSAPD--LQTAAVFAFLYSFTASFCG 1250  
QY 554 LTVNLFTTASWLSMLQYFSIPRYGF-----TALQHN-----EPLGQ 589  
Db 1251 VMQPYSLFPRF--WTMWRVSPPTYTIEFVSLILLHDREVNCGSTSENVQPMVGQCGG 1308  
QY 590 NFGPGLNATG-----NNP---CNATCT--GEELYVQKIDLSFPMGLMKH-----VA 632  
Db 1309 FMKPFIDFEGCKLHINNTYTVCAVCMYVQDDFLAQE--NMSYHHMRNRFPMVFCVN 1366  
QY 633 LACMIVIFLTIAVLKLF 650  
Db 1367 IAAMFVGFY-LTYIKKIW 1383

## RESULT 14

US-09-270-767-43468  
; Sequence 43468, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: Patent Ver. 2.0  
; SEQ ID NO 43468  
; LENGTH: 520  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-43468

Query Match 14.4%; Score 483.5; DB 4; Length 520;  
Best Local Similarity 29.4%; Pred. No. 4.4e-41;  
Matches 128; Conservative 78; Mismatches 162; Indels 67; Gaps 8;  
QY 60 EKEILSNINGIMKP-GLNALIGPTGGKSLDVLAAKDPGLSGDVLINGAPRANFX 118  
Db 90 EROILDSVGSFVPCVLAIMGSSGKTTLLDCLSGQRIID--SGSVFLNREPLTYKWR 147  
QY 119 CNGYVVDVVMGTLTVRENIQFSALRLATMTNHEKERNIRVIOELGDKVADSKV 178  
Db 148 RRIQVYLOEIEFPQTLTLETVVYVYALLRLPESMPAEKMRQVDHILEALELCCQQTKE 207  
QY 179 GTQPIRGVSGGERKRTSIGMELTDPGSLDEPTGLDSTANAVALLKMSK-QGRT 237  
Db 208 GDVNLNGLSGGEKRRKNINACELLTNPLMLLDEPTGLDSHSAISLMKVLKRYAQLEQKT 267  
QY 238 IIFSIIHQPRYSIFKLFDSTLLASGRLMFPGQOALGYESAGYCEAYNPPADFLD- 296  
Db 268 IIVSVIQPSSQMHMDKLLHOGKATAYGDVQNYLRHEDIGVITIKRYNADFLBQ 327  
QY 297 -----LINDSTAVALNR-----BEDFKATEIIPSKQDKPLIEKLAIEYV 337  
Db 328 LKSHPIREKLFYAKESHGNYLNRCITSSHHNVSVGAKGKKQADSLIDDIINNY 387  
QY 338 N-----SSFYETKALEQLSGGK 357  
Db 388 NQSRNHHQYENLHRTSNGCRVEDEEAQHLWCADQSNSFSCASSDCHSYSGSG 447

Qy 358 KKKITVFKKISYTSFCHQLRWVSKRSPKULLGNPQ-ASIAQIYTVVLGLVIGAIYFGL 416  
Db 448 PCHSADDDWLSTYTSFHTQFCVLSRNFRF--AKPRMLSKLWMPFOTIGLALMAGAIWFOQL 505  
Qy 417 KNDSTGIONRAGVLP 431  
Db 506 PRTEEFHLHDLQGMF 520

## RESULT 15

US-09-248-796A-20770  
; Sequence 20770, Application US/09248796A  
; Patent No. 6741137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 20770  
; LENGTH: 300  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-20770

Query Match 13.8%; Score 463; DB 4; Length 300;

Best Local Similarity 41.7%; Pred. No. 2.4e-39; Matches 106; Conservative 49; Mismatches 87; Indels 12; Gaps 6;

Qy 61 KEILSNINGIMKPG-LNAILGPTGGKSSLDVLAARKDP--SGLSGVLLINGAPRPAN- 116  
Db 27 KTLIDDIYGSVAGEMLAIMGSGGCKSTLNLVLAIRTSPRSTLEGIFINNERATLNK 86  
Qy 117 FKCSGYVQDDVNGTLTVRENLOPSAALRLATTMTNHEKNERINRVIOELGDKVADS 176  
Db 87 IKQLSSVVEQEDSLIGSLVLETVDYSAQF--AGIDKAHKKELVSKTKISGLEQAML 143  
Qy 177 KVGTFIRGVSGGKERTSIGMELITDPBILFLDEPTGLDSTANAVLLIKRMSK--QG 235  
Db 144 KIGTPIQKISGGQKRVSIASQIITSPIFLDEPTSGLDVSAAREVISTIKTIKAKREN 203  
Qy 236 RTIIFSIHQPRYSIFKLFDSTLTLLASGRIMFHPAQEALGFESAGYHCEAYNPDAPFL 295  
Db 204 MWIISIHQPSYTFELFDKVMFLSKGRIVNGAIVSNVKKYFNSIGHTMPPIYNPAEYVL 263  
Qy 296 DIIN---GDSTAV 305  
Db 264 DLINTDFOGDSSTVL 277

Search completed: June 6, 2005, 09:45:19  
Job time : 45 sec

**THIS PAGE LEFT BLANK**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

## OM protein - protein search, using bw model

Run on: June 6, 2005, 09:43:50 ; Search time 141 Seconds

(without alignments)  
1605.818 Million cell updates/sec

Title: US-09-961-086-1

Perfect score: 3352

Sequence: 1 MSSSVVEFIPVSGQNTNGF.....MIVFLTAVLKLFKKYS 655

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1465611 seqs, 345679903 residues

Total number of hits satisfying chosen parameters: 1465611

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3352	100.0	655	10	US-09-961-086-1
2	3352	100.0	655	15	US-10-405-806-13
3	3346	99.8	655	9	US-09-981-353-35
4	3346	99.8	655	14	US-10-120-687-61
5	3346	99.8	655	15	US-10-405-806-2
6	3346	99.8	655	17	US-10-874-706-24
7	3338	99.6	655	9	US-09-866-866A-27
8	3331	99.4	655	9	US-09-866-866A-10
9	3331	99.4	655	14	US-10-090-455-5
10	3053.5	91.1	604	9	US-09-745-763-197
11	2757	82.2	657	9	US-09-866-866A-14
12	835.5	24.9	1049	15	US-10-369-493-1520
13	812	24.2	663	13	US-10-108-605-245

14	812	24.2	687	17	US-10-732-923-1785	Sequence 1785, Ap
15	812	24.2	687	17	US-10-732-923-1786	Sequence 1786, Ap
16	812	24.2	696	17	US-10-732-923-1784	Sequence 1784, Ap
17	808	24.1	687	17	US-10-732-923-1775	Sequence 1775, Ap
18	806	24.0	687	17	US-10-732-923-1791	Sequence 1791, Ap
19	805	24.0	1084	15	US-10-424-599-242078	Sequence 242078, A
20	805	24.0	1101	15	US-10-425-114-63125	Sequence 63125, A
21	795.5	23.7	676	15	US-10-369-493-3799	Sequence 3799, Ap
22	793	23.7	692	17	US-10-732-923-1781	Sequence 1781, Ap
23	790	23.6	692	17	US-10-732-923-1777	Sequence 1777, Ap
24	777.5	23.2	679	17	US-10-732-923-1588	Sequence 1588, Ap
25	773.5	23.1	695	15	US-10-424-599-176182	Sequence 176182, A
26	772	23.0	695	17	US-10-732-923-1760	Sequence 1760, Ap
27	772	23.0	679	17	US-10-732-923-1783	Sequence 1783, Ap
28	760	22.7	696	17	US-10-732-923-1592	Sequence 1592, Ap
29	760	22.7	706	17	US-10-732-923-1789	Sequence 1789, Ap
30	754.5	22.5	677	17	US-10-732-923-1794	Sequence 1794, Ap
31	751	22.4	669	17	US-10-732-923-1773	Sequence 1773, Ap
32	751	22.4	695	17	US-10-732-923-1778	Sequence 1778, Ap
33	746.5	22.3	819	15	US-10-425-114-54421	Sequence 54421, A
34	745.5	22.2	673	17	US-10-732-923-1779	Sequence 1779, Ap
35	744	22.2	669	17	US-10-732-923-1774	Sequence 1774, Ap
36	743.5	22.2	1528	17	US-10-732-923-1678	Sequence 1678, Ap
37	743	22.2	1328	17	US-10-732-923-1677	Sequence 1677, Ap
38	739	22.0	612	16	US-10-437-963-185375	Sequence 185375, A
39	736.5	22.0	709	17	US-10-732-923-1682	Sequence 1682, Ap
40	730.5	21.8	1520	17	US-10-732-923-1680	Sequence 1680, Ap
41	728.5	21.7	711	16	US-10-437-963-123860	Sequence 123860, A
42	727	21.7	832	15	US-10-267-502-303	Sequence 303, Ap
43	723.5	21.6	700	16	US-10-437-963-180018	Sequence 180018, A
44	716	21.4	610	17	US-10-732-923-1765	Sequence 1765, Ap
45	715.5	21.3	1393	17	US-10-732-923-1681	Sequence 1681, Ap

## ALIGNMENTS

```
RESULT 1
US-09-961-086-1
; Sequence 1, Application US/09961086
; Publication No. US20030036645A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
; APPLICANT: ROSS, Douglas D.
; APPLICANT: DOYLE, L. Austin
; APPLICANT: ABRUZZO, Lynne
; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
; FILE REFERENCE: EP19376-019
; CURRENT APPLICATION NUMBER: US/09/961,086
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/073,763
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/02577
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-086-1
```

Query Match 100.0%; Score 3352; DB 10; Length 655;  
Best Local Similarity 100.0%; Pred. No. 2.2e-288;  
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSSVVEFIPVSGQNTNGFPATASNDIKAFTEGAVLSFPHNICYRVKXKSGSLPCKRPPVE 60  
DB 1 MSSSVVEFIPVSGQNTNGFPATASNDIKAFTEGAVLSFPHNICYRVKXKSGSLPCKRPPVE 60  
QY 61 KEILSINIGIMKPGINALIGPTGGGKSLDLVLAARKDPSGLSGDVLINGAPRPANPKCN 120

Db 61 KEILSNIINGIMKGLNAILGPTGGKSSLLDVLAAKDPGSLGSDVLIINGAPPANPKCN 120  
Qy 121 SGYVDDVVMGTLTYRENIQFSAALRLATTMNHEKNERINRYIOELGLDKVADSKVGT 180  
Db 121 SGYVDDVVMGTLTYRENIQFSAALRLATTMNHEKNERINRYIOELGLDKVADSKVGT 180  
Qy 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240  
Db 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240  
Qy 241 SIHOPRYSIFKLPDSITLLASGRLMFHPGPAQOALGFESAGYHCEAYNNPADFPDIIING 300  
Db 241 SIHOPRYSIFKLPDSITLLASGRLMFHPGPAQOALGFESAGYHCEAYNNPADFPDIIING 300  
Qy 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360  
Db 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360  
Qy 361 ITVFKESITTSFCHOLRWVSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGKNDK 420  
Db 361 ITVFKESITTSFCHOLRWVSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGKNDK 420  
Qy 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFHEIYISGYRVSSYFLGKLLSDLP 480  
Db 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFHEIYISGYRVSSYFLGKLLSDLP 480  
Qy 481 MTMLPSIIFTCIYFMGLKPKADAFVMMFTLMMVAYSASSMALAIAAGOSVSVATLL 540  
Db 481 MTMLPSIIFTCIYFMGLKPKADAFVMMFTLMMVAYSASSMALAIAAGOSVSVATLL 540  
Qy 541 MTICFVPMIIFSGLVNLTITIASWLSWLOFYSIPRGFTALQHNEFLGQFCGLNATGN 600  
Db 541 MTICFVPMIIFSGLVNLTITIASWLSWLOFYSIPRGFTALQHNEFLGQFCGLNATGN 600  
Qy 601 NPCNVATCTGEEYLVKQIDLSFWMGLMKHVALACMIVIFLTIAYLKLFLLKXYS 655  
Db 601 NPCNVATCTGEEYLVKQIDLSFWMGLMKHVALACMIVIFLTIAYLKLFLLKXYS 655

## RESULT 2

US-10-405-806-13  
; Sequence 13, Application US/10405806  
; Publication No. US2003023262A1  
; GENERAL INFORMATION:  
; APPLICANT: KOMATANI, HIDEYA  
; APPLICANT: HARA, YOSHIKAZU  
; APPLICANT: KOTANI, HIDEHITO  
; APPLICANT: NAKAGAWA, RINAKO  
; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF  
; FILE REFERENCE: 234985USCONT  
; CURRENT APPLICATION NUMBER: US/10/405, 806  
; CURRENT FILING DATE: 2003-04-03  
; PRIOR APPLICATION NUMBER: PCT/JP01/08112  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: JP2000-303441  
; PRIOR FILING DATE: 2000-10-03  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 13  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: ABCG2 482mutant sequence  
US-10-405-806-13

Query Match 100.0%; Score 3352; DB 15; length 655;  
Best Local Similarity 100.0%; Pred. No. 2, 2e-288;  
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MSSSNVEVFIPIVQSGNTNGPPTASNDLKAFTBGAVLSFNICRYRYLKSGLFPCRPVE 60  
Db 1 MSSSNVEVFIPIVQSGNTNGPPTASNDLKAFTBGAVLSFNICRYRYLKSGLFPCRPVE 60

Qy 61 KEILSNIINGIMKGLNAILGPTGGKSSLLDVLAAKDPGSLGSDVLIINGAPPANPKCN 120  
Db 61 KEILSNIINGIMKGLNAILGPTGGKSSLLDVLAAKDPGSLGSDVLIINGAPPANPKCN 120  
Qy 121 SGYVDDVVMGTLTYRENIQFSAALRLATTMNHEKNERINRYIOELGLDKVADSKVGT 180  
Db 121 SGYVDDVVMGTLTYRENIQFSAALRLATTMNHEKNERINRYIOELGLDKVADSKVGT 180  
Qy 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240  
Db 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKMSKQGRITIF 240  
Qy 241 SIHOPRYSIFKLPDSITLLASGRLMFHPGPAQOALGFESAGYHCEAYNNPADFPDIIING 300  
Db 241 SIHOPRYSIFKLPDSITLLASGRLMFHPGPAQOALGFESAGYHCEAYNNPADFPDIIING 300  
Qy 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360  
Db 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360  
Qy 361 ITVFKESITTSFCHOLRWVSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGKNDK 420  
Db 361 ITVFKESITTSFCHOLRWVSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGKNDK 420  
Qy 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFHEIYISGYRVSSYFLGKLLSDLP 480  
Db 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFHEIYISGYRVSSYFLGKLLSDLP 480  
Qy 481 MTMLPSIIFTCIYFMGLKPKADAFVMMFTLMMVAYSASSMALAIAAGOSVSVATLL 540  
Db 481 MTMLPSIIFTCIYFMGLKPKADAFVMMFTLMMVAYSASSMALAIAAGOSVSVATLL 540  
Qy 541 MTICFVPMIIFSGLVNLTITIASWLSWLOFYSIPRGFTALQHNEFLGQFCGLNATGN 600  
Db 541 MTICFVPMIIFSGLVNLTITIASWLSWLOFYSIPRGFTALQHNEFLGQFCGLNATGN 600  
Qy 601 NPCNVATCTGEEYLVKQIDLSFWMGLMKHVALACMIVIFLTIAYLKLFLLKXYS 655  
Db 601 NPCNVATCTGEEYLVKQIDLSFWMGLMKHVALACMIVIFLTIAYLKLFLLKXYS 655

## RESULT 3

US-09-981-353-35  
; Sequence 35, Application US/09981353  
; Patent No. US20020160382A1  
; GENERAL INFORMATION:  
; APPLICANT: Lasek, Amy W.  
; APPLICANT: Jones, David A.  
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
; FILE REFERENCE: PA-0038 US  
; CURRENT APPLICATION NUMBER: US/09/981, 353  
; CURRENT FILING DATE: 2001-10-11  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 35  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1  
US-09-981-353-35

Query Match 99.8%; Score 3346; DB 9; length 655;  
Best Local Similarity 99.8%; Pred. No. 7, 5e-288;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 MSSSNVEVFIPIVQSGNTNGPPTASNDLKAFTBGAVLSFNICRYRYLKSGLFPCRPVE 60  
Db 1 MSSSNVEVFIPIVQSGNTNGPPTASNDLKAFTBGAVLSFNICRYRYLKSGLFPCRPVE 60  
Qy 61 KEILSNIINGIMKGLNAILGPTGGKSSLLDVLAAKDPGSLGSDVLIINGAPPANPKCN 120

```
Db      61 KEILSININGIMKPGNALILGPTGGKSSLLDYLAAKDBSGSLGVDLINGAPRPAKFKN 120
Qy      121 SGYVODDVVNGTLTVRENLOFSAALRLATTTMTHNEKNERINRVIOELGLDYADSKVGT 180
Db      121 SGYVODDVVNGTLTVRENLOFSAALRLATTTMTHNEKNERINRVIOELGLDYADSKVGT 180
Qy      181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLRKMSKOGRTIIF 240
Db      181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLRKMSKOGRTIIF 240
Qy      241 SIHOPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db      241 SIHOPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Qy      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETIVNSSFYKETKAEHLHOLSGGEKKKK 360
Db      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETIVNSSFYKETKAEHLHOLSGGEKKKK 360
Qy      361 ITVPKEISYTTSFCHQLRWVSKRSFKNLGNPOASIAQIITVVGLVIGALYFGLKXDS 420
Db      361 ITVPKEISYTTSFCHQLRWVSKRSFKNLGNPOASIAQIITVVGLVIGALYFGLKXDS 420
Qy      421 TGIORAGVLFPLTTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 480
Db      421 TGIORAGVLFPLTTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 480
Qy      481 MTMLPSIITFTCIYFMGLGKPKADAFVMMFTLMVAVASASSMALAIAAGOSVSVATLL 540
Db      481 MTMLPSIITFTCIYFMGLGKPKADAFVMMFTLMVAVASASSMALAIAAGOSVSVATLL 540
Qy      541 MTICVPMIIFSGGLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONPCGLNATGN 600
Db      541 MTICVPMIIFSGGLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONPCGLNATGN 600
Qy      601 NPCNYATCTGGEYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
Db      601 NPCNYATCTGGEYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
```

```
RESULT 4
US-10-120-687-61
; Sequence 61, Application US/10120687
; Publication No. US20030082155A1
; GENERAL INFORMATION:
; APPLICANT: Massachusetts General Hospital
; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating
; FILE REFERENCE: 3284/1235B
; CURRENT APPLICATION NUMBER: US/10/120,687
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: US60/169082
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: US 09/963,875
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/215109
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: US 60/238880
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/731261
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-120-687-61
```

Query Match 99.8%; Score 3346; DB 14; Length 655;  
Best Local Similarity 99.8%; Pred. No. 7.5e-288;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MSSNVEVEFIPVSOQNTNGFPATASNDLKAFTEGAVLSEFHNICYRWKLSKSGFLPCRKPYE 60
Db      1 MSSNVEVEFIPVSOQNTNGFPATASNDLKAFTEGAVLSEFHNICYRWKLSKSGFLPCRKPYE 60
Qy      61 KEILSININGIMKPGNALILGPTGGKSSLLDYLAAKDBSGSLGVDLINGAPRPAKFKN 120
Db      61 KEILSININGIMKPGNALILGPTGGKSSLLDYLAAKDBSGSLGVDLINGAPRPAKFKN 120
Qy      121 SGYVODDVVNGTLTVRENLOFSAALRLATTTMTHNEKNERINRVIOELGLDYADSKVGT 180
Db      121 SGYVODDVVNGTLTVRENLOFSAALRLATTTMTHNEKNERINRVIOELGLDYADSKVGT 180
Qy      181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLRKMSKOGRTIIF 240
Db      181 QIRGVSGGERKRTSIGMELITDPSILFLDEPTTGDSSTANAVLLLRKMSKOGRTIIF 240
Qy      241 SIHOPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db      241 SIHOPRYSIFKLFDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Qy      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETIVNSSFYKETKAEHLHOLSGGEKKKK 360
Db      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETIVNSSFYKETKAEHLHOLSGGEKKKK 360
Qy      361 ITVPKEISYTTSFCHQLRWVSKRSFKNLGNPOASIAQIITVVGLVIGALYFGLKXDS 420
Db      361 ITVPKEISYTTSFCHQLRWVSKRSFKNLGNPOASIAQIITVVGLVIGALYFGLKXDS 420
Qy      421 TGIORAGVLFPLTTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 480
Db      421 TGIORAGVLFPLTTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 480
Qy      481 MTMLPSIITFTCIYFMGLGKPKADAFVMMFTLMVAVASASSMALAIAAGOSVSVATLL 540
Db      481 MTMLPSIITFTCIYFMGLGKPKADAFVMMFTLMVAVASASSMALAIAAGOSVSVATLL 540
Qy      541 MTICVPMIIFSGGLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONPCGLNATGN 600
Db      541 MTICVPMIIFSGGLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONPCGLNATGN 600
Qy      601 NPCNYATCTGGEYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
Db      601 NPCNYATCTGGEYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
```

```
RESULT 5
US-10-405-806-2
; Sequence 2, Application US/10405806
; Publication No. US2003023262A1
; GENERAL INFORMATION:
; APPLICANT: KOMATANI, HIDEYA
; APPLICANT: HARA, YOSHIKAZU
; APPLICANT: KOTANI, HIDEHIITO
; APPLICANT: NAKAGAWA, RINKO
; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF
; FILE REFERENCE: 234985USOCNT
; CURRENT APPLICATION NUMBER: US/10/405,806
; PRIOR FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: PCT/JP01/08112
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: JP2000-303441
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-405-806-2
```

Query Match 99.8%; Score 3346; DB 15; Length 655;  
Best Local Similarity 99.8%; Pred. No. 7.5e-288;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY      1 MSSNVEVFIPIVSOQNTNGFPATASNDLKAFTGAVLSFHNICVRVLSKSGFLPCRKPYE 60
      1 MSSNVEVFIPIVSOQNTNGFPATASNDLKAFTGAVLSFHNICVRVLSKSGFLPCRKPYE 60
Db      1 MSSNVEVFIPIVSOQNTNGFPATASNDLKAFTGAVLSFHNICVRVLSKSGFLPCRKPYE 60
QY      61 KEILSININGIMKPGLNAILGPTGGKSSLLDVLAAKDPGSLGVDLINGAPPANPKCN 120
      61 KEILSININGIMKPGLNAILGPTGGKSSLLDVLAAKDPGSLGVDLINGAPPANPKCN 120
Db      61 KEILSININGIMKPGLNAILGPTGGKSSLLDVLAAKDPGSLGVDLINGAPPANPKCN 120
QY      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
Db      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
QY      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
Db      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
QY      241 SIHQPRYSIFKLPDLSLTLLASGRLMFHPGPAQELGYFESAGYHCEAVNNPADPFLDIING 300
      241 SIHQPRYSIFKLPDLSLTLLASGRLMFHPGPAQELGYFESAGYHCEAVNNPADPFLDIING 300
Db      241 SIHQPRYSIFKLPDLSLTLLASGRLMFHPGPAQELGYFESAGYHCEAVNNPADPFLDIING 300
QY      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETVNSSEFKETAEHLQLSGGEKXK 360
      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETVNSSEFKETAEHLQLSGGEKXK 360
Db      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETVNSSEFKETAEHLQLSGGEKXK 360
QY      361 ITVFKKISTYTSFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKLKND 420
      361 ITVFKKISTYTSFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKLKND 420
Db      361 ITVFKKISTYTSFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKLKND 420
QY      421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVEKKLFHEHYISGYRVSSYFLGKLSDDL 480
      421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVEKKLFHEHYISGYRVSSYFLGKLSDDL 480
Db      421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVEKKLFHEHYISGYRVSSYFLGKLSDDL 480
QY      481 MTMLPSIIFTCIYFPMGLKPKADAFVMMFTLMVAVASSMALALAAQSVSVATLL 540
      481 MTMLPSIIFTCIYFPMGLKPKADAFVMMFTLMVAVASSMALALAAQSVSVATLL 540
Db      481 MTMLPSIIFTCIYFPMGLKPKADAFVMMFTLMVAVASSMALALAAQSVSVATLL 540
QY      541 MTICVFPMIIFSGLLVNLTTIASWLSMLOYFSIPRYGFTALQHNEFLGQFCGLNATGN 600
      541 MTICVFPMIIFSGLLVNLTTIASWLSMLOYFSIPRYGFTALQHNEFLGQFCGLNATGN 600
Db      541 MTICVFPMIIFSGLLVNLTTIASWLSMLOYFSIPRYGFTALQHNEFLGQFCGLNATGN 600
QY      601 NPCNVAATCGEEXLVKQIDLSPWGLMKHVALACMIVIFLTAYLKLFLKXYS 655
      601 NPCNVAATCGEEXLVKQIDLSPWGLMKHVALACMIVIFLTAYLKLFLKXYS 655
Db      601 NPCNVAATCGEEXLVKQIDLSPWGLMKHVALACMIVIFLTAYLKLFLKXYS 655

RESULT 6
US-10-874-706-24
; Sequence 24, Application US/10874706
; Publication No. US20050048610A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: LAB, Preeti
; APPLICANT: YANG, Junming
; APPLICANT: YUE, Henry
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: TANG, Y. Tom
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: AZIMZAI, Yaida
; APPLICANT: LU, Dying Anna M.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: PATTERSON, Chandra
; TITLE OF INVENTION: HUMAN TRANSPORT PROTEINS
; FILE REFERENCE: PF-0709 PCT
; CURRENT APPLICATION NUMBER: US/10/874,706
; FILE REFERENCE: 2004-06-24
; PRIOR APPLICATION NUMBER: US/10/009,328
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/139,923; 60/148,177; 60/149,357; 60/162,287
; PRIOR FILING DATE: 1999-06-17; 1999-08-10; 1999-08-18; 1999-10-28
; NUMBER OF SEQ ID NOS: 86
```

```
; SOFTWARE: PERL Program
; SEQ ID NO 24
; LENGTH: 655
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 5517972CD1
US-10-874-706-24

Query Match      99.8%; Score 3346; DB 17; Length 655;
Best Local Similarity 99.8%; Pred. No. 7,5e-288;
Matches 654; Conservative 0; Mismatch 1; Indels 0; Gaps 0;

QY      1 MSSNVEVFIPIVSOQNTNGFPATASNDLKAFTGAVLSFHNICVRVLSKSGFLPCRKPYE 60
      1 MSSNVEVFIPIVSOQNTNGFPATASNDLKAFTGAVLSFHNICVRVLSKSGFLPCRKPYE 60
Db      1 MSSNVEVFIPIVSOQNTNGFPATASNDLKAFTGAVLSFHNICVRVLSKSGFLPCRKPYE 60
QY      61 KEILSININGIMKPGLNAILGPTGGKSSLLDVLAAKDPGSLGVDLINGAPPANPKCN 120
      61 KEILSININGIMKPGLNAILGPTGGKSSLLDVLAAKDPGSLGVDLINGAPPANPKCN 120
Db      61 KEILSININGIMKPGLNAILGPTGGKSSLLDVLAAKDPGSLGVDLINGAPPANPKCN 120
QY      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
Db      121 SGVVODDVMGTLTYRENLOPSAALRLATMTNHNKRNIRNVIOELGLDKVADSKVGT 180
QY      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
Db      181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGSDSTANAVLLLLKRMKSQGRITIF 240
QY      241 SIHQPRYSIFKLPDLSLTLLASGRLMFHPGPAQELGYFESAGYHCEAVNNPADPFLDIING 300
      241 SIHQPRYSIFKLPDLSLTLLASGRLMFHPGPAQELGYFESAGYHCEAVNNPADPFLDIING 300
Db      241 SIHQPRYSIFKLPDLSLTLLASGRLMFHPGPAQELGYFESAGYHCEAVNNPADPFLDIING 300
QY      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETVNSSEFKETAEHLQLSGGEKXK 360
      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETVNSSEFKETAEHLQLSGGEKXK 360
Db      301 DSTAVALNREEDPKATEIIEPSKODKPLIEKLAETVNSSEFKETAEHLQLSGGEKXK 360
QY      361 ITVFKKISTYTSFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKLKND 420
      361 ITVFKKISTYTSFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKLKND 420
Db      361 ITVFKKISTYTSFCHOLRWVSKRSFKNLGNPQASIAQIIVTVVLGIVGAIYFGKLKND 420
QY      421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVEKKLFHEHYISGYRVSSYFLGKLSDDL 480
      421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVEKKLFHEHYISGYRVSSYFLGKLSDDL 480
Db      421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVEKKLFHEHYISGYRVSSYFLGKLSDDL 480
QY      481 MTMLPSIIFTCIYFPMGLKPKADAFVMMFTLMVAVASSMALALAAQSVSVATLL 540
      481 MTMLPSIIFTCIYFPMGLKPKADAFVMMFTLMVAVASSMALALAAQSVSVATLL 540
Db      481 MTMLPSIIFTCIYFPMGLKPKADAFVMMFTLMVAVASSMALALAAQSVSVATLL 540
QY      541 MTICVFPMIIFSGLLVNLTTIASWLSMLOYFSIPRYGFTALQHNEFLGQFCGLNATGN 600
      541 MTICVFPMIIFSGLLVNLTTIASWLSMLOYFSIPRYGFTALQHNEFLGQFCGLNATGN 600
Db      541 MTICVFPMIIFSGLLVNLTTIASWLSMLOYFSIPRYGFTALQHNEFLGQFCGLNATGN 600
QY      601 NPCNVAATCGEEXLVKQIDLSPWGLMKHVALACMIVIFLTAYLKLFLKXYS 655
      601 NPCNVAATCGEEXLVKQIDLSPWGLMKHVALACMIVIFLTAYLKLFLKXYS 655
Db      601 NPCNVAATCGEEXLVKQIDLSPWGLMKHVALACMIVIFLTAYLKLFLKXYS 655

RESULT 7
US-09-866-866A-27
; Sequence 27, Application US/0986686A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
```

;; PRIOR FILING DATE: 1999-05-27  
;; PRIOR APPLICATION NUMBER: 60/086,988  
;; PRIOR FILING DATE: 1998-05-28  
;; NUMBER OF SEQ ID NOS: 27  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 27  
;; LENGTH: 655  
;; TYPE: PRT  
;; ORGANISM: Homo sapien  
US-09-866-866A-27

Query Match 99.6%; Score 3338; DB 9; Length 655;  
Best Local Similarity 99.7%; Pred. No. 3.8e-287;  
Matches 653; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 MSSNVEVFIPIVSGQNTGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPVE 60
DB 1 MSSNVEVFIPIVSGQNTGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPVE 60
QY 61 KEILSININGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
DB 61 KEILSININGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
QY 121 SGVYVDDVVMGTLIVRENIQFSALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180
DB 121 SGVYVDDVVMGTLIVRENIQFSALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGSDSTANAVLLLRMSKQGTITIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGSDSTANAVLLLRMSKQGTITIF 240
QY 241 SIHQPRYSIFKLFDSLITLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
DB 241 SIHQPRYSIFKLFDSLITLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
QY 301 DSTAVALNEEDFPKATEIIEPSKODKPLIEKLAELIVNSSFYKETAEIHLQSGGEKKKK 360
DB 301 DSTAVALNEEDFPKATEIIEPSKODKPLIEKLAELIVNSSFYKETAEIHLQSGGEKKKK 360
QY 361 ITVFKEISYTTSFCHQLRWKRSRFRNLLGNPOASIAQIIVTVVGLVIGAIYFGKANDS 420
DB 361 ITVFKEISYTTSFCHQLRWKRSRFRNLLGNPOASIAQIIVTVVGLVIGAIYFGKANDS 420
QY 421 TGIQRAGVLPFLTNQCFSSVSAVELFVVEKKLFHEIISGYRVSSYFLGKLSDDL 480
DB 421 TGIQRAGVLPFLTNQCFSSVSAVELFVVEKKLFHEIISGYRVSSYFLGKLSDDL 480
QY 481 MTMLPSIIFTCIVYFNLGLKPKADAFVVMFTLMVAVASASSMALAIAAGQSVSVATLL 540
DB 481 MTMLPSIIFTCIVYFNLGLKPKADAFVVMFTLMVAVASASSMALAIAAGQSVSVATLL 540
QY 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOFYSIPRYGFTALQHNELGONFCPLNATGN 600
DB 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOFYSIPRYGFTALQHNELGONFCPLNATGN 600
QY 601 NPCNVAATCGEERYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
DB 601 NPCNVAATCGEERYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
```

RESULT 8  
US-09-866-866A-10

;; Sequence 10, Application US/09866866A  
;; Patent No. US20020102244A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Sorrentino, Brian  
;; APPLICANT: Schuetz, John  
;; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells  
;; FILE REFERENCE: 1340-1-02CICP2  
;; CURRENT APPLICATION NUMBER: US/09/866,866A  
;; PRIOR FILING DATE: 2001-08-30  
;; PRIOR APPLICATION NUMBER: 09/584,586  
;; PRIOR FILING DATE: 2000-05-31

;; PRIOR APPLICATION NUMBER: PCT/US99/11825  
;; PRIOR FILING DATE: 1999-05-27  
;; PRIOR APPLICATION NUMBER: 60/086,988  
;; PRIOR FILING DATE: 1998-05-28  
;; NUMBER OF SEQ ID NOS: 27  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 10  
;; LENGTH: 655  
;; TYPE: PRT  
;; ORGANISM: Homo sapien  
US-09-866-866A-10

Query Match 99.4%; Score 3331; DB 9; Length 655;  
Best Local Similarity 99.4%; Pred. No. 1.6e-286;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MSSNVEVFIPIVSGQNTGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPVE 60
DB 1 MSSNVEVFIPIVSGQNTGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCRKPVE 60
QY 61 KEILSININGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
DB 61 KEILSININGIMKPGNALIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRPAFKCN 120
QY 121 SGVYVDDVVMGTLIVRENIQFSALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180
DB 121 SGVYVDDVVMGTLIVRENIQFSALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGSDSTANAVLLLRMSKQGTITIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGSDSTANAVLLLRMSKQGTITIF 240
QY 241 SIHQPRYSIFKLFDSLITLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
DB 241 SIHQPRYSIFKLFDSLITLASGRLMFHGPAQALGFESAGYHCEAYNNPADPFLDIING 300
QY 301 DSTAVALNEEDFPKATEIIEPSKODKPLIEKLAELIVNSSFYKETAEIHLQSGGEKKKK 360
DB 301 DSTAVALNEEDFPKATEIIEPSKODKPLIEKLAELIVNSSFYKETAEIHLQSGGEKKKK 360
QY 361 ITVFKEISYTTSFCHQLRWKRSRFRNLLGNPOASIAQIIVTVVGLVIGAIYFGKANDS 420
DB 361 ITVFKEISYTTSFCHQLRWKRSRFRNLLGNPOASIAQIIVTVVGLVIGAIYFGKANDS 420
QY 421 TGIQRAGVLPFLTNQCFSSVSAVELFVVEKKLFHEIISGYRVSSYFLGKLSDDL 480
DB 421 TGIQRAGVLPFLTNQCFSSVSAVELFVVEKKLFHEIISGYRVSSYFLGKLSDDL 480
QY 481 MTMLPSIIFTCIVYFNLGLKPKADAFVVMFTLMVAVASASSMALAIAAGQSVSVATLL 540
DB 481 MTMLPSIIFTCIVYFNLGLKPKADAFVVMFTLMVAVASASSMALAIAAGQSVSVATLL 540
QY 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOFYSIPRYGFTALQHNELGONFCPLNATGN 600
DB 541 MTICVFPMWIFSGLLVNLTTIASWLSWLOFYSIPRYGFTALQHNELGONFCPLNATGN 600
QY 601 NPCNVAATCGEERYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
DB 601 NPCNVAATCGEERYLVKQIGIDLSPMGLMKHVALACMIVIFLTIAVLKLLFLKKYS 655
```

RESULT 9

US-10-090-455-5  
;; Sequence 5, Application US/10090455  
;; Publication No. US20030027259A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Chen, Hongyun  
;; APPLICANT: Le Bihan, Stephane  
;; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF  
;; FILE REFERENCE: 100103.406  
;; CURRENT APPLICATION NUMBER: US/10/090,455  
;; PRIOR FILING DATE: 2002-03-01  
;; NUMBER OF SEQ ID NOS: 17

SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 655  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-090-455-5

Query Match 99.4%; Score 3331; DB 14; Length 655;  
Best Local Similarity 99.4%; Pred. No. 1,6e-286;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MSSNVEVFIPVQSGNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCKRPVE 60
DB 1 MSSNVEVFIPVQSGNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCKRPVE 60
QY 61 KEILSNINGIMKGLNAIIGPTGGKSSLLDVLAARDPSGLSGDVLINAPRPAFKCN 120
DB 61 KEILSNINGIMKGLNAIIGPTGGKSSLLDVLAARDPSGLSGDVLINAPRPAFKCN 120
QY 121 SGVVODDVMGTLTVRENIFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180
DB 121 SGVVODDVMGTLTVRENIFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANA VLLLLKMSKQGRTIIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANA VLLLLKMSKQGRTIIF 240
QY 241 SIHQPRYSIFKLPDSITLLASGRLMFHGPAQELGYFESAGYHCEAYNNPADFFLDING 300
DB 241 SIHQPRYSIFKLPDSITLLASGRLMFHGPAQELGYFESAGYHCEAYNNPADFFLDING 300
QY 301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAIEYVNSFYKETRAELHQLSGEKKKK 360
DB 301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAIEYVNSFYKETRAELHQLSGEKKKK 360
QY 361 ITVFKEISYTTSFCHQLRWVSKSPKULLGNPOASIAQIIVTVVLGIVGAIYFGKNDK 420
DB 361 ITVFKEISYTTSFCHQLRWVSKSPKULLGNPOASIAQIIVTVVLGIVGAIYFGKNDK 420
QY 421 TGIQNAGVLFLLTNOCFSSVSAVELFVVEKKLFHEYISGYRRSSYFLGLSLDLPL 480
DB 421 TGIQNAGVLFLLTNOCFSSVSAVELFVVEKKLFHEYISGYRRSSYFLGLSLDLPL 480
QY 481 MTMLPSIIFTCIYFMLGLKPKADAFVMMFTLMVAAYASSMALAIAGQSVVAATLL 540
DB 481 MTMLPSIIFTCIYFMLGLKPKADAFVMMFTLMVAAYASSMALAIAGQSVVAATLL 540
QY 541 MTICVFPMIFSGLVNLTITIASWLSWLYFSIPRYGFTALQHNELQNFCEGLNATGN 600
DB 541 MTICVFPMIFSGLVNLTITIASWLSWLYFSIPRYGFTALQHNELQNFCEGLNATGN 600
QY 601 NPCNVAITCGEEYLVKQIDLSPMGLMKHVALACMIVIFLITAYLKLFKKYS 655
DB 601 NPCNVAITCGEEYLVKQIDLSPMGLMKHVALACMIVIFLITAYLKLFKKYS 655
```

## RESULT 10

US-09-745-763-197

Sequence 197, Application US/09745763

Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John M.

Lavallie, Edward R.

Collins-Racie, Lisa A.

Evans, Cheryl

Merberg, David

Treacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

ENCODING THEM

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/745,763  
FILING DATE: 18-Jun-2000  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 197:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 604 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 197:  
US-09-745-763-197

Query Match 91.1%; Score 3053.5; DB 9; Length 604;  
Best Local Similarity 91.9%; Pred. No. 6,3e-262;  
Matches 602; Conservative 0; Mismatches 2; Indels 51; Gaps 1;

```
QY 1 MSSNVEVFIPVQSGNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCKRPVE 60
DB 1 MSSNVEVFIPVQSGNTNGFPATASNDLKAFTGAVLSFHNICRYVKLSGFLPCKRPVE 60
QY 61 KEILSNINGIMKGLNAIIGPTGGKSSLLDVLAARDPSGLSGDVLINAPRPAFKCN 120
DB 61 KEILSNINGIMKGLNAIIGPTGGKSSLLDVLAARDPSGLSGDVLINAPRPAFKCN 120
QY 121 SGVVODDVMGTLTVRENIFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180
DB 121 SGVVODDVMGTLTVRENIFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANA VLLLLKMSKQGRTIIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANA VLLLLKMSKQGRTIIF 240
QY 241 SIHQPRYSIFKLPDSITLLASGRLMFHGPAQELGYFESAGYHCEAYNNPADFFLDING 300
DB 241 SIHQPRYSIFKLPDSITLLASGRLMFHGPAQELGYFESAGYHCEAYNNPADFFLDING 300
QY 301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAIEYVNSFYKETRAELHQLSGEKKKK 360
DB 301 DSTAVVALNREDEPKATEIIEPSKODKPLIEKLAIEYVNSFYKETRAELHQLSGEKKKK 360
QY 361 ITVFKEISYTTSFCHQLRWVSKSPKULLGNPOASIAQIIVTVVLGIVGAIYFGKNDK 420
DB 361 ITVFKEISYTTSFCHQLRWVSKSPKULLGNPOASIAQIIVTVVLGIVGAIYFGKNDK 420
QY 421 TGIQNAGVLFLLTNOCFSSVSAVELFVVEKKLFHEYISGYRRSSYFLGLSLDLPL 480
DB 421 TGIQNAGVLFLLTNOCFSSVSAVELFVVEKKLFHEYISGYRRSSYFLGLSLDLPL 480
QY 481 MTMLPSIIFTCIYFMLGLKPKADAFVMMFTLMVAAYASSMALAIAGQSVVAATLL 540
DB 481 MTMLPSIIFTCIYFMLGLKPKADAFVMMFTLMVAAYASSMALAIAGQSVVAATLL 540
QY 541 MTICVFPMIFSGLVNLTITIASWLSWLYFSIPRYGFTALQHNELQNFCEGLNATGN 600
DB 541 MTICVFPMIFSGLVNLTITIASWLSWLYFSIPRYGFTALQHNELQNFCEGLNATGN 600
```

Db 490 MTICFVFMNIFSGLLVNLTTIASWLSWLOFYSIPRYGFTALQHNELGONFCPLNATGN 549  
QY 601 NNCNATCTGGEYLVKQIDLSPWGIMKRNHVALACMIVIFLTAIVKLLFLKYS 655  
Db 550 NNCNATCTGGEYLVKQIDLSPWGIMKRNHVALACMIVIFLTAIVKLLFLKYS 604

## RESULT 11

US-09-866-866A-14  
/ Sequence 14, Application US/09866666A  
/ Patent No. US20020102244A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Sorrentino, Brian  
/ APPLICANT: Schuetz, John  
/ TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells  
/ FILE REFERENCE: 1340-1-021CIP2  
/ CURRENT APPLICATION NUMBER: US/09/866,866A  
/ CURRENT FILING DATE: 2001-08-30  
/ PRIOR APPLICATION NUMBER: 09/584,586  
/ PRIOR FILING DATE: 2000-05-31  
/ PRIOR APPLICATION NUMBER: PCT/US99/11825  
/ PRIOR FILING DATE: 1999-05-27  
/ PRIOR APPLICATION NUMBER: 60/086,988  
/ PRIOR FILING DATE: 1998-05-28  
/ NUMBER OF SEQ ID NOS: 27  
/ SOFTWARE: PatentIn version 3.0  
/ SEQ ID NO 14  
/ LENGTH: 657  
/ TYPE: PRF  
/ ORGANISM: Mus musculus  
US-09-866-866A-14

Query Match 82.2%; Score 2757; DB 9; Length 657;  
Best Local Similarity 81.5%; Pred. No. 1,6e-235;

Matches 536; Conservative 51; Mismatches 67; Indels 4; Gaps 3;

QY 1 MSSNVEVEFIPVSGQNTGFPATASNDLKAFTEGAVLSFHNICRYVKLSGFLPCCKPYE 60  
Db 1 MSSNDHVLVPMQSRNNNLPRMNSKAVTLTLAGDVLSPHHTYRKVKSGL-VAKTVE 59  
QY 61 KEILSNINGIMKPG-LNAILGPTGGKSSLDVLAARKDPSGLSGDVLINGARPAANFKCN 120  
Db 60 KEILDINDIMKPG-LNAILGPTGGKSSLDVLAARKDPSGLSGDVLINGARPAANFKCN 119  
QY 121 SGYVQVDDVVMGTLTYRENLQFSAALRLATTTMNEKNERINRVIOELGLDKVADSKVT 180  
Db 120 SGYVQVDDVVMGTLTYRENLQFSAALRLPTTMKNEKNERINTYIKELGLEKVDASKVT 179  
QY 181 QTRIVSGGERKRTSISGMELITDPSILFLDEPTTGIDSTANAVLLLLKRMKQGRITIF 240  
Db 180 QTRIVSGGERKRTSISGMELITDPSILFLDEPTTGIDSTANAVLLLLKRMKQGRITIF 239  
QY 241 SIHOPRYSIFKLFDSTLLASGRLMFHPGPAQALGFYSAGYHCEAYNNPADFFLDI 300  
Db 240 SIHOPRYSIFKLFDSTLLASGRLMFHPGPAQALGFYSAGYHCEAYNNPADFFLDI 299  
QY 301 DSTAVANLREE-DFKATELIEPSKODKPLIEKLAETIYVNSSEYKETAELHQLSGEKKK 359  
Db 300 DSSAVMLNREEQDNANKTEEPSKCEKPYIENLSEFYNSAIYGETKALDQLPGAQEKK 359  
QY 360 KITVKEISYTTSPFCHQLRWVSKSPKLNLPQASIAQIITVVUIGVIGAIYGLKND 419  
Db 360 GTSAREPEYVTSFCHQLRWVSKSPKLNLPQASIAQIITVVUIGVIGAIYGLKND 419  
QY 420 STGIQNRAGVLFPLTTNQCFSSVASVELFVVEKKLFIHEIYSGYRVSSYFLGKLLSDL 479  
Db 420 AAGQNRAGVLFPLTTNQCFSSVASVELFVVEKKLFIHEIYSGYRVSSYFLGKMSDDL 479  
QY 480 PNTMLPSIITFCIYVFMGLKPKKADAFVMMFTLMMVAVSASSMALAIAGQSVSVATL 539  
Db 480 PNRFLPSVIFCIYVFMGLKKTVDAAFIMFTLLMVAVTASSMALAIATGQSVSVATL 539  
QY 540 LMTICFVFMNIFSGLLVNLTTIASWLSWLOFYSIPRYGFTALQHNELGONFCPLNATG 599

Db 540 LMTIAFVFMNIFSGLLVNLRTIGPWLSQLQYISIPRYGFTALQYNEFLGQFCFGNATND 599  
QY 600 NRPC--NATCTGGEYLVKQIDLSPWGIMKRNHVALACMIVIFLTAIVKLLFLKYS 655  
Db 600 NSTCVNSYAICTGNEYLINQGIELSPWGLMKRNHVALACMIIIFLTAIVKLLFLKYS 657

## RESULT 12

US-10-369-493-1520  
/ Sequence 1520, Application US/10369493  
/ Publication No. US20030233675A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Cao, Yongwei  
/ APPLICANT: Hinkle, Gregory J.  
/ APPLICANT: Slater, Steven C.  
/ APPLICANT: Goldman, Barry S.  
/ APPLICANT: Chen, Xianfeng  
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
/ FILE REFERENCE: 38-10(52052)B  
/ CURRENT APPLICATION NUMBER: US/10/369,493  
/ CURRENT FILING DATE: 2003-02-28  
/ PRIOR APPLICATION NUMBER: US 60/360,039  
/ PRIOR FILING DATE: 2002-02-21  
/ NUMBER OF SEQ ID NOS: 47374  
/ SEQ ID NO 1520  
/ LENGTH: 1049  
/ TYPE: PRF  
/ ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-1520

Query Match 24.9%; Score 835.5; DB 15; Length 1049;  
Best Local Similarity 30.5%; Pred. No. 1,6e-64;

Matches 222; Conservative 134; Mismatches 257; Indels 115; Gaps 18;

QY 1 MSSNVEVEFIPVSGQNTGFPATASNDLKAFTEGAVLSFHNICRYVKLSGFLPCCKPYE 60  
Db 355 LQSSNSPRLP-DEAVNNFLONEDDTL-----ATLSENTITYSVPSINS-----DGYE 402  
QY 61 KEILSNINGIMKPG-LNAILGPTGGKSSLDVLAARKDPSGLSGDVLINGARPAANFKCN 118  
Db 403 ETVLNEISIGIVAPGQILAIMGSSGAGKTTLLDILMKRKTGHVSSIKXNGISMBRKSS 462  
QY 119 CNSGYVQVDDVVMGTLTYRENLQFSAALRLATTTMNEKNERINRVIOELGLDKVADSKV 178  
Db 463 KIIQVDDDFLLPTLTYFEVJNLSALRLPVALLSFEAKKAVYVLELRIIDIKR11 522  
QY 179 GTRFIRGVSGGERKRTSISGMELITDPSILFLDEPTTGIDSTANAVLLLLKRMKQGRITIF 237  
Db 523 GNEFDRGISGGERKRTSISGMELITDPSILFLDEPTTGIDSTANAVLLLLKRMKQGRITIF 582  
QY 238 IIFSHPRYSIFKLFDSTLLASGRLMFHPGPAQALGFYSAGYHCEAYNNPADFFLDI 297  
Db 583 LVLSTHQRNSNIFLYFDKVLVLSKEMRYSGAKKAVSELRNEGICPNYNTIADYLDI 642  
QY 238 -----INDSTAV 305  
Db 643 TFEAPQGRRRIRNISDLAAGTDNDIDNTIHTFTSSDGTQREMAHLAHLDEIRS 702  
QY 366 ALNREDEKATE-----IIEPSKODKPLIEKLAETIYVNSSEYKETAELHQLSGEKKK 360  
Db 703 LIRDEDEVEGDRGATEIDINTLHLDK---FKDSYVAELSQELVEVLSGDEBSN 758  
QY 361 IT--AFKEISYTTSPFCHQLRWVSKSPKLNLPQASIAQIITVVUIGVIGAIYGLKND 418  
Db 759 VLNQGLPFGQGSAGLQGLSTLNSRSFNNMVRNPKLLGNVLTLLSLFGLTYNNYN 818  
QY 419 DSTGIQNRAGVLFPLTTNQCFSSVASVELFVVEKKLFIHEIYSGYRVSSYFLGKLLSDL 478  
Db 819 DISGRQNMGLFFFLTYFGFVTFGLSSFALERIFKERSNNYSPLAAYYSKIMSEV 878  
QY 479 LPMTMLPSIITFCIYVFMGLKPKKADAFVMMFTLMMVAVSASSMALAIAGQSVSVAT 538



```

Db      879 VPLRVVPPILSLIVPMGTGLNMKDNAFFKCIIGILLIFNLGISLEILLITIGIFEDLNNSI 938
Qy      539 LMTICVFMWIFSGLLV--NLTTASWLSWLOYESIPRYGFTALOHNEF----- 566
Db      939 ILTVLVLLGSLPSGLFINTKNTTVA--FKYLNFSVFYYAESLLINEVKTLMLKERE 996
Qy      587 LQGNF-CPGIATGNPCNYATCTGEEYLVKQGI--DLSPWGLMKHVALACMIVIFLTI 643
Db      997 YGNIIEVPG-----ATLSTRGFVVQNLVFDIK-----ILALFNVVFLIM 1036
Qy      644 AYTKLFL 651
Db      1037 GYLLAKWI 1044

RESULT 13
US-10-108-605-245
; Sequence 245, Application US/10108605
; Publication No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stem, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108, 605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 245
; LENGTH: 663
; TYPE: PR1
; ORGANISM: Drosophila melanogaster
US-10-108-605-245

Query Match      24.2%; Score 812; DB 13; Length 663;
Best Local Similarity 32.1%; Pred. No. 9.5e-63;
Matches 210; Conservative 134; Mismatches 251; Indels 60; Gaps 17;

Qy      5 NVEVFIPVSGQNTNGPATNSDLKAFTEGAVLSFHNICRVYKLSGFLPCRPVEKEIL 64
Db      50 NMDIFGAVNQ-----PGSGMRQLVNRTRGLFCNERHI-----PAPR---KHL 89
Qy      65 SNINGIMKPG-LNAIIGPTGGKSSLLDVLAAKDPGSL---SGDVLINGAPRPA-NFK 118
Db      90 KNYGVAVPEELAVWSSGAGKTTLLNALAFR-SFGIOVSPSGMRLNGQVDAKEMQ 148
Qy      119 CNSGYVODDVVNGTLTVRENLOFSAALRLATTTNHEKNERINRYIOELGLDKVADSKV 178
Db      149 ARCAVYQODDLFGISLTAREHLIFQAMVMPRHLYRQVARVDQVIOELSLSKQHTII 208
Qy      179 GTQ-FIRGVSGGRRKRTSIGMELITDPSILFLDEPTGLDSSSTANAVLLLKMSKQGT 237
Db      209 GVPGRVYSGSGGRRKRLAFASEALTDPLICDEPTSGDLSFPAHSVQVLYKLSQSGKT 268
Qy      238 IIFSIHOPRSIFKLFDLSLTLLASGRLMFHPGPAQELAGYESAGYHCEAVNPNADFLDI 297
Db      269 VILTIHOPSELELFDKILMLMEGRVAFGLTSEAVDFSYGAOCPTNYNPNADFLYQV 328
Qy      298 INGSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIYVNSFFYKETAELHQLSGEK 357
Db      329 L-----AVVPGREIBER-----DRIKICDNPAISKVAR-DMEQLLATKN 367
Qy      358 KKKIIVFKEISY--TSFCHQLRWVSKSFKNLGNPQASIAQIIVTVVLGLVIGAIYFG 415
Db      368 LEKPLQEPENGYYTKATFWQFRAVLWRMSLVLEKEPLLVKVRLIQTTVAIILIGLIFLG 427
```

```

Qy      416 LKNDSTGIONRAGVLEFLTTNOCFSSVSA-VLEFVVEKGLFIHEIYISGYRVSSYFLGKL 474
Db      428 QQLTVGWNMINCAIILPLTNMFQVNPATINVFTESELPEFMEARSRLRYRCOTYFLGKT 487
Qy      475 LSDLMLMTMPSIIFICIYIFMGLKPKADAPFVMMFTLMVAVYSSAMALAAQOSV 534
Db      488 IAE-LPLFLTVPLVFAIYIPMIGRAGVLAHFENCALVTLVANVSTSGYLLISCASSST 546
Qy      535 SVATLMTICFVPMI.FSGLVNLTITASWLSWLOYESIPRYGFTALOHNEFLGQNFPCPG 594
Db      547 SMLASVGPVYIIFLFLFGCFNLNGSVPVYIKLSTLSMFRYANEGLLINQWADVE--RG 604
Qy      595 -LNAVGNPCNYATCTGEEYLVKQGI.DLSPWGLMKHVALACMIVIFLTIAYIKL 648
Db      605 EISCTSSN-----TTCSSGKVILETLNFSADPLDYGVGLALIVSFRVLYALAL 655

RESULT 14
US-10-732-923-1785
; Sequence 1785, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgercon, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 1785
; LENGTH: 687
; TYPE: PR1
; ORGANISM: Drosophila melanogaster
US-10-732-923-1785

Query Match      24.2%; Score 812; DB 17; Length 687;
Best Local Similarity 32.1%; Pred. No. 1e-62;
Matches 210; Conservative 134; Mismatches 251; Indels 60; Gaps 17;

Qy      5 NVEVFIPVSGQNTNGPATNSDLKAFTEGAVLSFHNICRVYKLSGFLPCRPVEKEIL 64
Db      74 NMDIFGAVNQ-----PGSGMRQLVNRTRGLFCNERHI-----PAPR---KHL 113
Qy      65 SNINGIMKPG-LNAIIGPTGGKSSLLDVLAAKDPGSL---SGDVLINGAPRPA-NFK 118
Db      114 KNYGVAVPEELAVWSSGAGKTTLLNALAFR-SFGIOVSPSGMRLNGQVDAKEMQ 172
Qy      119 CNSGYVODDVVNGTLTVRENLOFSAALRLATTTNHEKNERINRYIOELGLDKVADSKV 178
Db      173 ARCAVYQODDLFGISLTAREHLIFQAMVMPRHLYRQVARVDQVIOELSLSKQHTII 232
Qy      179 GTQ-FIRGVSGGRRKRTSIGMELITDPSILFLDEPTGLDSSSTANAVLLLKMSKQGT 237
Db      233 GVPGRVYSGSGGRRKRLAFASEALTDPLICDEPTSGDLSFPAHSVQVLYKLSQSGKT 292
Qy      238 IIFSIHOPRSIFKLFDLSLTLLASGRLMFHPGPAQELAGYESAGYHCEAVNPNADFLDI 297
Db      293 VILTIHOPSELELFDKILMLMEGRVAFGLTSEAVDFSYGAOCPTNYNPNADFLYQV 352
Qy      298 INGSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIYVNSFFYKETAELHQLSGEK 357
Db      353 L-----AVVPGREIBER-----DRIKICDNPAISKVAR-DMEQLLATKN 391
Qy      358 KKKIIVFKEISY--TSFCHQLRWVSKSFKNLGNPQASIAQIIVTVVLGLVIGAIYFG 415
Db      392 LEKPLQEPENGYYTKATFWQFRAVLWRMSLVLEKEPLLVKVRLIQTTVAIILIGLIFLG 451
Qy      416 LKNDSTGIONRAGVLEFLTTNOCFSSVSA-VLEFVVEKGLFIHEIYISGYRVSSYFLGKL 474
Db      452 QQLTVGWNMINCAIILPLTNMTFQVNPATINVFTESELPEFMEARSRLRYRCOTYFLGKT 511
```

```

Oy 475 :SDDLPMTPMLPSIIIFPCIIYVFMIGLXPKADAFVMMFTLMMVYSSSMALAIAGGSV 534
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 IAE-LPFLTVLVVFTALYIPMIGLAGVLIHFNCIALVTLVANNVSTISFGYLLISCHSST 570
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 535 SVATLMTICPFYFMIMFSGLLVNLTTIASWLSWQFSIIPRYGFTALQHNHFSGQNFPCG 594
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 571 SNAISVGPVITLFPFLIFGFGFLNNGSVPYILKLKSLYSLSWRVYNEGILLINQMDVFE--PG 628
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 595 -LNATGNPCNATYCTGBEYLVKQCIDLSPMGLMKRHVALACMIVIFLTAYAKL 648
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 639 EISCSSN-----TTCPSSGKVLLETINFSAADLPDYGVGALITVSRVLAIAL 679
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      571  SMLASVGPVPIIPFLFGSPFLNSSGVPYLLKMLSYLSMFRANEGCLLNQMDVE--PG 628
      :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY      555  -LMTGNNPCMYATTTGEEYLVKQSIDLSPKQIMKHVALACMIYFLLTAYIKL 648
      :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db      629  EISCTSSN----TTCPSGSKVILETLNFSADPLPDYGLATLIVSFRVLAAYLAL 679
      :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

```

RESULT 15
US-10-732-923-1786
: Sequence 1786, Application US/10732923
: Publication No. US20050108791A1
: GENERAL INFORMATION:
: APPLICANT: Edgerton, Michael D
: TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPE
: FILE REFERENCE: 38-15152796/C
: CURRENT APPLICATION NUMBER: US/10/732,923
: CURRENT FILING DATE: 2003-12-10
: PRIOR APPLICATION NUMBER: 10/310,154
: PRIOR FILING DATE: 2002-12-04
: NUMBER OF SEQ ID NOS: 24149
: SEQ ID NO 1786
: LENGTH: 687
: TYPE: RRT
: ORGANISM: Drosophila melanogaster
US-10-732-923-1786

```

```

Search completed: June 6, 2005, 09:57:41
Job time : 144 secs

```

Query Match	24.2%	Score 812;	DB 17;	Length 687;
Best Local Similarity	32.1%	Pred No. 1e-62;		
Matches 210; Conservative	133;	Mismatches 251;	Indels 60;	Gaps 17;

```

QY 5 NVEVFIPIVSOQGNITNFPAPASNDLKAFTGCAVLSFNNI CVRYLKGSGFLPCRPEVEKEI 64
Db 74 NMDITGAVNQ-----PGSGRQLVNRKTGCLCNEHNI-----PAPF---KHLI 113
QY 65 SNINGIMKRG-LNALIGPYGGKSSLLDVLAARKDPSG----SGDVLINGAPRAP-NPK 118
Db 114 KNVCGVAYPGBELLAVMGSSGACKTTLNLNLAFR-SFGQIQVSPSGRLLNGQVDAKEMQ 172
QY 119 CNSGYVODDDVVMGTLTVARENQFSAALFLATTMTNHEKNERINRYIOELGDKVADSKY 178
Db 173 ARCAVVOODDLFIQSLTAREHLIFQAMVMPRHLLTYRQVARVDQVIOELSLSKOHTII 232
QY 179 GTQ-FIRVSGGERKRTSGMELLDPSLFLDEPTGTGDSSTANNVLLIKRMSQGR 237
Db 233 GVPGRKSLSGGERRLRAPASALTDPPLLICDEPSTGDSFPAHSHVVOLLKLSQGRKT 292
QY 238 IIFSIIHQPRYSIFKLFDSLTLLASGRLMHGPAQELAGYFESAGYCEAYNNPADPFLDI 297
Db 293 VILTIHQSSSELFELFDKILMAEGRVAFGLTGPSEAVDFSFYVGAQCPNNYMPADPYVOY 352
QY 298 INGSSTAVALNREEDFKATELIEPESKODKPLIEKLAETIVNSSPYKETAELHOLSGEK 357
Db 353 L-----AVVPGREIESR-----DRIAKICDNFASIKVAR-DMEQLLATKN 391
QY 358 KKKITVFEKIST--TSFCHQLRWYSKSKSFKULGNPQASIAOIYTVVLGLVITAIIRG 415
Db 392 LEKPLEOPENGTYTAKATWFMOPRALWMBRLSVLEKPLLVKRLIOTTTVALILGIFLG 451
QY 416 LKNDSTGIONRAGVFLPLTNOCSSPVSA-VELFVVEKFLTHEYISGYRVSSYFLGKL 474
Db 452 QOLTOGVANNINGAIFLPLTNMTFQVRFATINFTSELEFVKEARSRLYKCDTHYLGKT 511
QY 475 LSDLPLMTLPSIIFTCIYFMLGLKPKADAEFVMFTLMMVAYSSAMALAIAGQSVV 534
Db 512 IAE-LPLFLTVLVEFTALAIYPMIGLRAGVLHFNCLALTVLVANNSTSGYILSCSSST 570
QY 535 SVATILMTITCEVFMNIPSGLLVNLTTIASWLSWLOFSIPRYGFTALQHNEFLGONFCG 594

```

THIS PAGE LEFT BLANK

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2005, 09:27:54 ; Search time 44 Seconds  
(without alignments)  
1432.318 Million cell updates/sec

Title: US-09-961-086-1

Perfect score: 3352  
Sequence: 1 MSSNVEVFIPVSGQNTNGF.....MIVFLTAYLKLFLKKYS 655

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Limiting filter 45 summaries

Database : PIR 79:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2849.5	85.0	656	2 JC7860	brain multidrug re
2	835.5	24.9	1049	1 S19421	ATP-dependent perm
3	812	24.2	687	1 FYFWM	white protein - fr
4	800.5	23.9	737	2 T46101	ABC transporter-11
5	774	23.1	646	2 C86441	probable ABC trans
6	767.5	22.9	687	2 D96553	hypothetical prote
7	746	22.3	649	2 A84509	probable ABC trans
8	745	22.2	725	2 T47652	ABC transporter-11
9	741.5	22.1	739	2 T45891	ABC transporter-11
10	739	22.0	678	2 H96552	hypothetical prote
11	726.5	21.7	708	2 T47650	ABC transporter-11
12	721	21.5	635	2 T08934	hypothetical prote
13	720.5	21.5	755	2 G84791	probable ABC trans
14	717.5	21.4	740	1 T0567	probable ATP-bind
15	712.5	21.3	609	2 E96742	probable ABC trans
16	710	21.2	638	2 G02068	white homolog - hu
17	708	21.1	646	2 JC7777	ATP binding caset
18	707.5	21.1	547	2 T31543	hypothetical prote
19	707.5	21.1	520	2 T47648	ABC transporter-11
20	677	20.2	759	2 B88474	protein C05D10.3
21	676	20.2	725	2 C84423	probable ABC trans
22	669	20.0	659	2 E86313	hypothetical prote
23	664	19.8	608	2 T34391	hypothetical prote
24	660	19.7	662	2 T47649	ABC transporter-11
25	658.5	19.6	1294	2 S77690	probable membrane
26	657.5	19.6	1450	2 T45888	ABC transporter-11
27	656.5	19.6	590	2 B96573	protein F12M16.17
28	649.5	19.4	633	2 T19189	hypothetical prote
29	645.5	19.3	577	2 T04229	ABC-type transp

30	639	19.1	658	2 T31958	hypothetical prote
31	636.5	19.0	639	2 G88839	protein C10C6.5 [1
32	636.5	19.0	695	2 T21109	hypothetical prote
33	627.5	18.7	610	2 T19333	hypothetical prote
34	612.5	18.3	1501	2 S50992	SMO2 protein - yea
35	603	18.3	1530	2 S52239	bretefelin A resist
36	612	18.0	1530	2 T52010	hypothetical prote
37	602	18.0	1511	2 A53151	pleiotropic drug r
38	594	17.7	1564	2 S55517	probable transp
39	593	17.7	1443	2 T02491	probable ABC trans
40	590	17.6	1469	2 H96622	probable ABC trans
41	585	17.5	1333	2 S63403	probable membrane
42	580.5	17.3	1420	2 T02644	ABC-type transp
43	580.5	17.3	1529	2 S69688	hypothetical prote
44	572	17.1	1413	2 G84790	probable ABC trans
45	564	16.8	1466	2 T30566	ATP-binding caset

#### ALIGNMENTS

##### RESULT 1

JC7860  
brain multidrug resistance protein, BMDP - pig  
C/Species: Sus scrofa domestica (domestic pig)  
C/Date: 18-Nov-2002 #sequence\_revision 18-Nov-2002 #ext\_change 09-Jul-2004  
C/Accession: JC7860

R/Isenblatter, T.; Galia, H.J.

Biochem. Biophys. Res. Commun. 293, 1273-1278, 2002

A/Title: A new multidrug resistance protein at the blood-brain barrier.

A/Reference number: JC7860; PMID:12050127; PMID:12054514

A/Accession: JC7860

A/Molecule type: mRNA

A/Residues: 1-656 <EIS>

A/Cross-references: UNIPROT:Q8M1B3; GB:AJ420927

A/Experimental source: brain

A/Comment: This protein, a new transport protein of the ATP-binding cassette (ABC) superfamily, is expressed in the brain and participates in drug transport across the blood-brain barrier.

A/Genes: bmdp

Query Match	Score	DB 2	Length
Best Local Similarity	84.3%	Pred. No. 6.9e-192	
Matches	553	Conservative	44; Mismatches 58; Indels 1; Gaps 1;
QY	1	MSSNVEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICRYVKLKGFLPCRKYVE	60
DB	1	MSSNVEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICRYVKLKGFLPCRKYVE	60
QY	61	KEILSNINIMKPGNALILGPTGGKSSLDVLAARKDPSGLSGVLTNGARPNPKCN	120
DB	61	KEILSNINIMKPGNALILGPTGGKSSLDVLAARKDPSGLSGVLTNGARPNPKCN	120
QY	121	SGYVQDDVMGTLTVRENLQPSAALRLATWTNHEKNERIRRVIOELGLDKVADSKVCT	180
DB	121	SGYVQDDVMGTLTVRENLQPSAALRLATWTNHEKNERIRRVIOELGLDKVADSKVCT	180
QY	181	QPIRVSQGERKRTSIGMELITDPSILFLDEPTTGDSSTANAVILLLRMSKQRTIIF	240
DB	181	QPIRVSQGERKRTSIGMELITDPSILFLDEPTTGDSSTANAVILLLRMSKQRTIIF	240
QY	241	SIHQPRYSIFKFLFSDLTLLASGRLMFHPGPAQALGYFESAGHCHCEAYNNPAPFLDING	300
DB	241	SIHQPRYSIFKFLFSDLTLLASGRLMFHPGPAQALGYFESAGHCHCEAYNNPAPFLDING	300
QY	301	DSTAVALNR-EEDFPATEIIEPSKODKPLIEKLAIEYVNSPFYKTKAEHLQSGEKKK	359
DB	301	DSTAVALNR-EEDFPATEIIEPSKODKPLIEKLAIEYVNSPFYKTKAEHLQSGEKKK	359
QY	360	KITVEKEISYTTSPFCHQLRWVSKRSFKULLGNPQASIAQIIVTVVGLVGAIVGLKND	419
DB	360	KITVEKEISYTTSPFCHQLRWVSKRSFKULLGNPQASIAQIIVTVVGLVGAIVGLKND	419

**RESULT 2**

ATP-dependent permease ADP1 precursor - Yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: protein YCR011c; protein YCR105  
C:Species: *Saccharomyces cerevisiae*  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C:Accession: S19421; S40914  
R:Goffeau, A.; Purnelle, B.; Skala, J.  
submitted to the Protein Sequence Database, March 1992  
A:Reference number: S19420  
A:Accession: S19421  
A:Molecule type: DNA  
A:Residues: 1-1049 <GOR>  
A:Cross-references: UNIPROT:P25371; EMBL:X59720; NID:g1907116; PIDN:CAA42328.1; PID:g1907116; F:Skala, B.; Goffeau, A.  
Yeast 7, 867-872, 1991  
A>Title: The product of the YCR105 gene located on the chromosome III from *Saccharomyces cerevisiae*.  
A:Reference number: S40914; PMID:92160395; PMID:1769009  
A:Accession: S40914  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-1049 <PUR>  
R:Skala, J.; Purnelle, B.; Goffeau, A.  
Yeast 8, 409-417, 1992  
A>Title: The complete sequence of a 10.8 kb segment distal of SUF2 on the right arm of chromosome III.  
K genes.  
A:Reference number: S25353; PMID:92327849; PMID:1626432  
A:Contents: annotation  
C:Genetics:  
A:Gene: SGD:ADP1; MIPS:YCR011c  
A:Cross-references: SGD:S0000604; MIPS:YCR011c  
A:Map position: 3R  
C:Superfamily: ATP-dependent permease ADP1; ATP-binding cassette homology C/Keywords: ATP; glycoprotein; nucleotide binding; F-loop; transmembrane protein F:1-55/Domain: signal sequence #status predicted <SIG>  
F:26-1049/Product: ATP-dependent permease ADP1 #status predicted <EXT>  
F:26-324/Domain: extracellular #status predicted <EXT>  
F:325-341/Domain: transmembrane #status predicted <TM1>  
F:406-607/Domain: ATP-binding cassette homology <ABC>  
F:423-430/Region: nucleotide-binding motif A (P-loop)  
F:550-557/Region: nucleotide-binding motif B  
F:794-810/Domain: transmembrane #status predicted <TM2>  
F:829-845/Domain: transmembrane #status predicted <TM3>  
F:878-894/Domain: transmembrane #status predicted <TM4>  
F:909-925/Domain: transmembrane #status predicted <TM5>  
F:938-954/Domain: transmembrane #status predicted <TM6>  
F:1025-1041/Domain: transmembrane #status predicted <TM7>  
F:50\_114\_165\_221/Binding site: carboxylate (Asn) (covalent) #status predicted <TM7>  
F:423/Binding site: ATP (lys) #status predicted

```

Db 355 LGSSKSPRLP-DEDAVNNFQJNEDPTL-----ATLSFNITVSVPSINS-----DGVE 402
QY 61 KEILSNINGIMKPG-LNAILGPTGGGSSLLDVLAAKNDGSLGVDVLINARPP-ANFK 118
Db 403 ETVINEISIGVYKPOQIIAIMGSSGAGTTLIDILAMKRKTGHVSGSIKVGISIMDRKFS 462
QY 119 CNSGVYVDDVVMGTLVTRNLOGSALTRATMTNHEKNERIRVIOEJGLDGVADSKV 178
Db 463 KIIFVQDDFLPTLVFETVYVNSALLRLPKALSFAKARVYKVUEELRIIDIKRII 522
QY 179 GTQFIRVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLRMSKQ-GRT 237
Db 523 GNEBDRGISGGEKRRVSIACELVTSPLVFLDEPTSGLDASNNANNVTECVRLSSDYNR 582
QY 238 IIFSIHOPRISIFKLFDPSLTLASGLRMFHGPAOALGVESAGYHEAEANRPADFLLDI 297
Db 583 LVLSIHOPRNSIFLPLKVLVLSKGEVWVSGNAKRVSEFLRNEGXIOPDNYNADVYLDI 642
QY 298 -----INGDSTAV 305
Db 643 TFEAPQOKRRRRINISDLEAGDTDNDITHOFTTSSDGTQREMAHLAAHRDEIRS 702
QY 306 ALNREEDFKATE-----IIEPSKODPLIEKLAELIYVNSFYKETKAELHQ-LSGGEKKK 360
Db 703 LIRBEDVEGIDGRGATEIJDANKLHDK-----YDSVYAAELSGOIEEVLBSGDESN 758
QY 361 IT--VFKEISYTSFCHQLRWVSKRSFKNLGNFOASIAOIYTVVLGVYIGALFYGLKN 418
Db 759 VLNGDLPTGOOSAGFLQOLSIINRSKSNMYRNPDKLLGVLLITLITLSLFLGTLYVWNS 818
QY 419 DSTGIORAGVLFLLTNQCFSSVSNVELFVVEKULFIHEYSIGSYVSVYFGLKLSDL 478
Db 819 DISGFQRMGJFEFLITLYFGVFTFTGSSALEREIIFIKRSNNYSPLAVYISKIMSEV 878
QY 479 LPMTMLSIIFTCVYFMLGLKPADAFVMMFLTMVAVYASASAMALAIAGOSVAVAT 538
Db 879 VPLKVVPPILSLIIVYPMTGANKMDNAFFKICIGLILFNIGISIELITLTGIFPEDLANSI 938
QY 539 LLMTCICEVPMVIFSGLLV--NLTTJASWLSWLOYSIIPRYGTALQHNEF----- 586
Db 939 ILSVLVLLGSLFSGFLINTKNITNVA--PKYLNKFSVFVYAYASLLINEVKTLMLKERK 996
QY 587 LGQNF-CBGLMATGNPCNATYTCGEEYLVKQGI--DLSPMGKMNHALACMIVIFLTI 643
Db 997 YGLNIEVVG-----ATLSTGFEVQNLVFDIK-----ILAEVNVFLIM 1036
QY 644 AYKLLFL 651
Db 1037 GYLLAKM 1044

```

```

Query Match      24.9%  Score 835.5;  DB 1;  Length 1049;
Best Local Similarity 30.5%  Pred. No. 2.3e+50;
Matches 222;  Conservative 134;  Mismatches 257;  Indels 115;  Gaps 18;

Oy      1  MSSSSVVEFIIVSQCNTNGFPAISNDKAFTGCAVLFPHNICYVKUKSGFLPCKRYVE 60
      :  ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

RESULT 3  
PYPFW  
white protein - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 31-Dec-1990 #sequence revision 17-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: S08635; S07263; S10240  
R:Repling, M.; Mount, S.M.  
Nucleic Acids Res. 19, 1633, 1990  
A:Title: Sequence of a cDNA from the *Drosophila melanogaster* white gene.  
A:Reference number: S08635; MUID:90221897; PMID:2109311  
A:Accession: S08635  
A:Molecule type: mRNA  
A:Residues: 1-687 <PEP>  
A:Cross-references: UNIPROT:P10090; EMBL:X51749; NID:g68825; PID:CAA16038.1; PUD:g68826  
R:O'Hare, K.; Murphy, C.; Lewis, R.; Rudin, G.M.  
J. Mol. Biol. 180, 437-455, 1984  
A:Title: DNA sequence of the white locus of *Drosophila melanogaster*.  
A:Reference number: S07263; MUID:85134865; PMID:6084717  
A:Accession: S07263  
A:Molecule type: DNA  
A:Residues: 1-74, 'LIFELIPHCRTAD', 30-334, 'ITLHLNSYPAWVSULPTTIRRTTTCWPLCPDGRSSPVIGSPRT'  
A:Cross-references: EMBL:X029374







## RESULT 7

A84509

probable ABC transporter [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #ext\_change 09-Jul-2004

C:Accession: A84509

R:Lin. X.; Kaul, S.; Rounleay, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.;

eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: A84509

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-649 &lt;STO&gt;

A:Cross-references: UNIPROT:Q9SIR6; GB:AE002093; NID:94558665; PIDN:AAD22683.1; GSPDB:GN

C:Genetics:

A:Gene: At2g13610

A:Map position: 2

C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-b

Query Match 22.3%; Score 746; DB 2; Length 649;

Best Local Similarity 33.5%; Pred. No. 2,2e-44;

Matches 203; Conservative 116; Mismatches 239; Indels 48; Gaps 18;

QY 61 KEILSNINGIMKP-GLNALIGPTGGKSSLLDVLAAKDPGSLGDPVLINGAP-REPANK 118

DB 60 KHVLRKQVCRAPWEIATVGPAGKSSLELTARLPQ--TGSVYVKKRPVDRANK 117

QY 119 CMSGYVODDVVMTGLTVENENQFSAALRLATMTHEKNERINRYIOGLGDKVADSKV 178

DB 118 KISGYVQDQTLFPLTVETELLFSAKLL--KLPADELRSRKSLVHLEGLAEVAATAV 175

QY 179 GTQPIRGVSGERKRISIMELITDPSILFDEPTTGLDSTANAVALLKMKSK-QGR 237

DB 176 GDDSVRGISGGERRRKRSIVIEVHDKVILDEPTSGLDSTSLIITDMLKMAERKGT 235

QY 238 IIFSIHQPRYSIFKLPDLSITLLASGLMFHQAQALGYESAGYCEAYNNPADFFLDI 297

DB 236 ILTTHQPRFRIYKQNSVLLANGSTLKQSGVDQGVYLRSGHLPRLHENVIEPAIS 295

QY 298 INGSATAVALNREDFKATEIIEPSKQKPLIEKLAETVNSFYKETVAELHOLSGEK 357

DB 296 I--ESTTKQQRLOESRRRAHVLP--QTLQEKSEDSQGES--KSGFTLLQQLFOQTR 348

QY 358 KKKI-TVPEISYTTSPCH---QLRWVSKSPKMLGNPQASINQIIVTVVLGIVIGAI 412

DB 349 VADVGTMNATEFTDRDFANSRLDETMTLTHRPSKNTFTKELPACTVOMLGSGLVGLI 408

QY 413 YGLKNDSTGIONRAGVLFELTNNQFSSVASVELFVEEKKLFIEHYISGYRVSSYFLG 472

DB 409 FHNLMDDKLGARERGLFAFILTLSTITLALPFLQERILMKETSSGSRVSSYANA 468

QY 473 KLLSDLEPMTLPSIIFTCIVYFMGLKPKADAFVYMMFTLMVAYSSAMALATAA--- 529

DB 469 NGLV-YLPFLILIALIFSTPYVMVLGNPSFMAFLFSLIIMLIYLTANSVVVCSALVP 527

QY 530 ----GQSVVSVATLMTICFVFMNIFSGLLVNLITLAWLSWLYQYSIRYGFYALQHNE 585

DB 528 NPIVGNSTVSG---VWGSFF---LPSGFTISNHEIPGWIMFMYISLKYPREGFLINE 580

QY 586 FLGQFPCPLANTGNPCNYATCTGEEYLVKQIDLSFWG---LWKNHVALACMIVFLT 642

DB 581 PAKSKMK---LEYGFGK---LVTEEDLKE---ERYGESRMRNVYIMCPVLLYVF 629

QY 643 IAYLKL 648

DB 630 ISYVIL 635

## RESULT 8

T47652

ABC transporter-like protein - Arabidopsis thaliana

N:Alternate names: protein T26112.10

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #ext\_change 09-Jul-2004

C:Accession: T47652

R:Montfort, A.; Casacuberta, E.; Pulidomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X

submitted to the Protein Sequence Database, February 2000

A:Reference number: Z24471

A:Accession: T47652

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-725 &lt;MON&gt;

A:Cross-references: UNIPROT:Q9M3D6; EMBL:AL132954

A:Experimental source: cultivar Columbia; BAC clone T26112

C:Genetics:

A:Map position: 3

A:Note: T26112.10

C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-b

Query Match 22.2%; Score 745; DB 2; Length 725;

Best Local Similarity 29.3%; Pred. No. 3e-44;

Matches 193; Conservative 130; Mismatches 274; Indels 62; Gaps 13;

QY 36 VLSFHNICRYVLRKSGFLPCRKVEKILSNINGIMKRP-LNALIGPTGGKSSLLDVA 94

DB 72 VLNFNLOYDVTLLRRRFGFSKQNGVKTLLDDVSGASDDIYAVLACASGAKSTLIDALA 131

QY 95 ARKDSGSLGDPVLINGAP--REPANKMSGYVODDVVMTGLTVENENQFSAALRLATMT 152

DB 122 GRVARGSLGSLGDPVLINGAP--REPANKMSGYVODDVVMTGLTVENENQFSAALRLATMT 191

QY 153 TNEKNERINRYIOGLGDKVADSKVGTQPIRGVSGERKRISIMELITDPSILFDEP 212

DB 192 SKSKIMBEREALIDOLGRNANANTVIGDEGHGVSGEERRRSIGDIIDHPIVFLDPR 251

QY 213 TTGLDSTANAVALLKMKSKQGRITIFSIHQPRYSIFKLPDLSITLLASGLMFHQAQAL 272

DB 252 TGLDSTANAVALLKMKSKQGRITIFSIHQPRYSIFKLPDLSITLLASGLMFHQAQAL 311

QY 273 ALGYESAGYCEAYNNPADFFLDI-----NGDSTAVALNREDFKATEIIEPS 322

DB 312 LPGFSDPGRPIPEKENISFPAIDLVRELEGSNEBTKALVDNEMKQOKISLIGAPOT 371

QY 333 ---KQDKPLIEKLAETVNSFYKETKAEHLQSGEKKKKITVFEKISYTTSPCHQLRW 379

DB 372 NKLDDRSLSLKEA---INASV---SRGL--VSGSSRSNPTSMETVSSYANPSLFEPR 423

QY 380 VSKRFPKMLGNPQASINQIIVTVVLGIVIGAIYGLKNDSTGIONRAGVLFELTNNQCF 439

DB 424 LAKRYMKWIRMPBELVGRIATVMVTCGLATVYKLDHTEPRGADERLTLFAFVPTMY 483

QY 440 SSVASVELFVEEKKLFIEHYISGYRVSSYFLGKLSLDLEPMTLPSIIFTCIVYFMGL 499

DB 484 CCLDNVPFIDGRYIFLEETTHNAVRTSSYVSHLV--LPDLAPSLVFSAITWTYGL 542

QY 500 KRKADAFVYMMFTLMVAYSSAMALATAAGSVSVATLMTICFVFMNIFSGLLVNLIT 559

DB 543 SCGLEGFYFVCLLIYASFSGSVVTFISGVVPNIMLCYMAVITVLAICLLLSGFVND 602

QY 560 TIASLWSWLYQYSIRYGFYALQHNEFLQNC-----PGLNAG----- 599

DB 603 RLPFYMTWFHYSILKYPEAVLINFPDPSCFVRGQVFPDSTLLGVSQSKVLLT 662

QY 600 ----NPNCAVATC--TGEELYLVKQGI--DLSFGLMKHVALACMIVFLTAYVLLP 650

DB 663 LSKSLRTITESTCLARTSGDLAAGGITQLSKMD-----CLMITPASGLFPRILF 712

## RESULT 9

T45891

ABC transporter-like protein - Arabidopsis thaliana

N/Alternate names: protein F4P12.210  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
C/Accession: T45891  
R/Biochecker: H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quefeler, F.; Salanoubat, M.  
submitted to the Protein Sequence Database, January 2000  
A/Reference number: 223016  
A/Accession: T45891  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-739 <BLO>  
A/Cross-references: UNIPROT:Q91FG8; EMBL:AL132966  
A/Experimental source: cultivar Columbia; BAC clone F4P12  
C/Genetics:  
A/Map position: 3  
A/Note: F4P12.210  
C/Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-h

Query Match 22.1%; Score 741.5; DB 2; Length 739;  
Best Local Similarity 28.3%; Pred. No. 5.5e-44;  
Matches 193; Conservative 123; Mismatches 256; Indels 111; Gaps 17;  
36 VLSFHNICVRVKKSGF--LPC--RKPE-----KEILSNINGMKDG-LNALIGPT 82  
87 VLSFKDLTVSVKTKKKKPPCCGNSPFGNDNEMNTKVLINGISGARSGEMMAVIGAS 146  
83 GGGKSLIDVLAARKDPSGLDVLINGADRPANF-KNSGYVVDVWGTITVRENLO 141  
147 GSGKSTLIDALARIKESLRGDTTLNGEVLSESLHVIAYVQDLDLFPMLTVEETLM 206  
142 FSAALRIATMTWHEKNERINRVIQELGDKVADSKVGTQFIRGSGEKRTSIGMELI 201  
207 FSAELFRLPSLSKKKKARVQALDQGLRMAAKTVIGDEGHRGVSGBERRRVSIGDII 266  
202 TDSILFLDEPTGDSSTANAVLLKRMKSGKRTIIFSIHORYSIFLFPDSTILLAS 261  
267 HDPIILFLDEPTGSDTSAYMVVKVQRIAGSGSIYVMSIHQPSYRILGLDLPLSR 326  
262 GRIMFHQPAOBALGYFESAGYHCEAVNPPADFLDINDGSTAVALNREDFKATEIEP 321  
327 GNIVYSSPTHLQOFEBEHPPIPENENKPEFALDLR-----ELSDS 369  
322 SKODKPLIEKLAETIYVNSFYKETAEHLQSGEKKKKITVEKEI----- 367  
370 PEGTKSLIVE-----FHKQWRK--QTSQSRNTNVSLEKDAISASISRGKIVSGA 417  
368 -----SYTTSFCHQLRWVSKSFKNLGNPQASIAQIIVVGLVGLVGIATYFGKND 419  
418 TNLRSFOTANPPWTEMLVIGKRSILNSRRQPELFGIRGAVLVGTMLIATTFWKLDNS 477  
420 STGIQNRAGVLFLLTNQCFSSSAVELFVVEKKLFHEYISGYRVSSYFLGKLSDDL 479  
478 PRGIQELGFFAFAMSTTFYTCAEAIPIVLOEKRIIFREITAYANAYRRSSSVLAHTIIS-I 536  
480 PMTMLPSIIFTCIYVFMGLKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATL 539  
537 PALIISAAFAASTFSFASVLAGSEGLPFFFTILTAFWAGSFEVTELSGVASHVIMIGFT 566  
540 LMTICFVFMIFSGGLVNLTTIASKLSWLOYSIPRIGFTALQNEFLGQNF--PGLNA 597  
597 VVAAILAYFLFLFGGFIISDRIDPLVYMWIMFHLISLVKPYEGVVLONEDEPTKCFVRDIOM 656  
598 TGNNPCNYA-----TC--TGEELYLVKQGI-DLSPMG-LMKNH 630  
657 FDSNPLQGVPTAVKISLLKSMGVLGINVTAETCVTTGIDILKQGGITBISKNCNM--- 713  
631 VALACMIVIFLTIA---YLLKLF 650  
714 -----ITVAMGFFFRVLV 726

RESULT 10  
H96552

hypothetical protein FSD21.8 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C/Accession: H96552  
R/Biochecker: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: H96552  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-678 <STO>  
A/Cross-references: UNIPROT:Q9C8J8; GB:AE005173; NID:gl0092361; PIDN:AA012770.1; GSPDB:(  
C/Genetics:  
A/Map position: 1  
C/Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-i

Query Match 22.0%; Score 739; DB 2; Length 678;  
Best Local Similarity 30.3%; Pred. No. 7.3e-44;  
Matches 186; Conservative 137; Mismatches 248; Indels 42; Gaps 13;  
61 KEILSNINGMKDG-LNALIGPTGGKSLIDVLAARKDPSG--LSGDTVLINGAPRAN 116  
28 KRILNNGGCGENRILAIIMPSSGSKTLLDLAARL--AGVNVSGKLVNGKKRRLD 85  
117 FKCSNGYVQDDVVMGTLTVRENLOPSAALRLATMTWHEKNERINRVIQELGDKVADS 176  
86 FGA-AAVYQEDVILGTLTVRESISYSAHLRLSKITREISIVERTITDMGLEBESDR 144  
177 KVGTOPIRGVSGERRRTSIGMELITDPSILFLDEPTGDSSTANAVLLKRMKSGKR 236  
145 TIGNMHLRGISGEKKRLSIALEVLTKPSILFLDEPTGSDSASAFVVOILNRIASSGK 204  
237 TIIFSIHORYSIFKLPDSITLTLASGRLMHGAQALGYFESAGYHCEAVNPPADFLD 236  
205 TVVSSIHQPSYFALFDLILSGEYVYFGAESATKFGGAGFPCCPERRNPSDFLR 264  
297 IINGD-----STAVANLR--EEDFKATEIIEPSKOKPLIEKLAETIYVNSFYKETAEHL 349  
265 CVNSDPDNTAALVESRRINDSFSILQHLETTNIDPL-DIPIAEIRITLVKPKFCSL 323  
350 HOLSGEKKKKITVF-----KEISYTSFCHQLRWVSKSFKNLGNPQASIAQIIVTV 403  
324 YAAASRARIGELIASIVGIYTERKKSGQTNMKQLRILITQGSFINMSRDLGAYMMRIAVYI 383  
404 VLGVLVIGALYFGKANDSTGIQNRAGVLFLLTNQCFSSSAVELFVVEKKLFHEYISGY 463  
384 VLSICVGSIFPFGVGRNHTNWSSTAACGFAGMTMISIGFOSFIEEMVCFRERLNGH 443  
464 YRVSSYFLGKLSDDLPMTMLPSIIFTCIYVFMGLKPKADAFVMMFTLMVAYSASGM 523  
444 YGVAIVTWSLBS-LPFTILMCLSTSSITTYNWRQSGSHFFYNCLDICALITYBSC 502  
524 ALAIAAGQSVSVATLMTIC---FVFMNIFSGGLVNLTTIASKLSWLOYSIPRYGFT 579  
503 MMMIAS-----VVPNFMGLMGVLAGYIGIVLVAGFRFPDLPVPMWRPVGSYINGAW 557  
580 ALQ--HNEFLGQFCPGLNATGNPCNATCTGEEYLVKQGIDLSFGWGLMKHVALACM 636  
558 ALQGAAYNEMIGVEY-----DSPLPVPRKKGELLITQTVGINBESSKWLDAVVM 609  
637 IVFELTAYLKL 649  
610 ILIGRIAPFALL 622



probable ABC transporter [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004

C/Accession: G84791

R./lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

eues, D.; Nlemaan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A./Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A./Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: G84791

A/Status: preliminary

A./Molecule type: DNA

A./Residues: 1-755 <STO>

A./Cross-references: UNIPROT:Q9ZUT0; GB:AE002093; NID:G4056489; PIDN:AAC98055.1; GSPDB:GN

C/Genetics:

A/Map position: 2

C/Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-b

Query Match

21.5%; Score 720.5; DB 2; Length 755;

Best Local Similarity 28.3%; Pred. No. 1,7e-42;

Matches 202; Conservative 133; Mismatches 282; Indels 99; Gaps 21;

OY 2 SSSNVEVFPVSGQNTNGPAPATASNDLKAFTEGA-VLSPHNICRYVKLKGFLP--CRK 57

DB 63 SSRALGIASPINBA-ASSFMSWASAPASSISSPVLSPFDLTVYVKIQKFNPLACRR 121

OY 58 PVE-----KEISNNGIKKPG-LNALIGPTGGKSLDVLAAKDPGSLGSDVLING 110

DB 122 SGNDSSVNTKILLNGISGEAREGEMAVLNASGSGKSTLLDALANRAKDSLGSTLLNG 181

OY 111 APRPAMP-KCNSGVVODDVMVMTGLTVREMLQPSAALRLATMTNHEKNERIRNIOELG 169

DB 182 EVLESSMOKISAYWODDLFPMILVTEETLMSAEERLRSLSKSKKKARQALDOLG 241

OY 170 LDRVADSKVGTOPRIGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLK 229

DB 242 LRSAAKTVIDEGHGVSGGERRRVSGNDIHDPIILFLDEPTSGDSTAYVIVKVLQ 301

OY 230 RMSKQRTTIFSHOPIYSIFKLPDSTLLASGRIMHGAQELGYFESAGHCEAYNN 289

DB 302 RIAQSGSIVIMS.IHQPYSRIMGLDOLIFLSKGTVSGSPTHLPQFSFKPIPIENEN 361

OY 290 PADFLDIINGDSTAVLANREEDPKATEIIEPSKODPLIE-----KLAETIYV-- 338

DB 362 KTEFALDLI-----RELVS-----TEGTPELVPHIKOMRAKQAPSYNNKKR 404

OY 339 ----SSFYKETKALHQ--LSGGEKKKKITVFEKI--SYTTSFCHOLRWVSKSFKNLLG 390

DB 405 NTWVSLKEAITASISRGKLVSGATNNNSNLPSPFQTPANPFMIEMIVIGKALINSRR 464

OY 391 NPQASIAQIIVTVVGLVIGAIYFGLKNDSTGIONRGVLPFLTNOCSVSANVELFV 450

DB 465 QPRLGRLCAVWVTGIIATMTFLNDSKGAQERLGFAPFAMSTTFYCAEAIPIVFLQ 524

OY 451 EKLFIHEYISGYRVSSYFLGKLISDLPLMTLPSIIFTCIYVFMGLPKADAFVVM 510

DB 525 ERITFEMETAYNARRSSIVLSOSIIS-IPALVLSAFAATTFFWAVGLGANGFFEFY 583

OY 511 FTLMVAVYSASMAALAAQOSVVSVALTLMTICFVPMIFSGLVNLTTIASVWLYQ 570

DB 584 FTILASFWAGSSFTVTLISGVIPNVLGFTVVAIIAFLFLFSGFPLSRDIIPIYVWLMFHY 643

OY 571 FSLPRGFTALQHNELPGON---FCPGLNATGNP-----C 603

DB 644 ISLVKPYEGVLQNEF--QNPTRCFAGVOLFQNSPLGEPNDVKVNLKMSGVLGTNV 701

OY 604 NVATFC--TGEVYVKOGI-DLSPMG-LMKNHVALACIVIFLTIA---YKLLF 650

DB 702 TAEFTVTTGIDILKQGGITDLSKNCILM-----ITVAMGFFRVL 742

RESULT 14

102567

probable ATP-binding cassette protein T16B24.1 - Arabidopsis thaliana

N/Alternate names: protein F12L6.1

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

C/Accession: T02567; T00545; G84816

R./lin, X.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,

submitted to the EMBL Data Library, August 1998

A./Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.

A./Reference number: Z14679

A/Accession: T02567

A/Status: translated from GB/EMBL/DBJ

A./Molecule type: DNA

A./Residues: 1-740 <ROW>

A./Cross-references: UNIPROT:O80946; EMBL:AC004697; NID:G3402671; PIDN:AAC28975.1; PID:G

A./Experimental source: cultivar Columbia

R./lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

eues, D.; Nlemaan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,

Nature 402, 761-768, 1999

A./Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A./Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: G84816

A/Status: preliminary

A./Molecule type: DNA

A./Residues: 1-362 <ROW>

A./Cross-references: EMBL:AC004218; NID:G3355463; PIDN:AAC27826.1; PID:G3355464

A./Experimental source: cultivar Columbia

R./lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

eues, D.; Nlemaan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,

Nature 402, 761-768, 1999

A./Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A./Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: G84816

A/Status: preliminary

A./Molecule type: DNA

A./Residues: 1-740 <STO>

A./Cross-references: GB:AE002093; NID:G3402672; PIDN:AAC28975.1; GSPDB:GN00139

C/Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-i

C/Keywords: ATP

F/110-310/Domain: ATP-binding cassette homology <ABC>

Query Match 21.4%; Score 717.5; DB 1; Length 740;

Best Local Similarity 28.0%; Pred. No. 2.6e-42;

Matches 193; Conservative 120; Mismatches 294; Indels 83; Gaps 16;

OY 26 NDLKAFTEGAVLSFHNICRYVKLK-----SGFLPCR-----KPYEKELSNING 69

DB 56 NDGYMRTVPVPSFDMLTVVSVRPKLDPRNLPPRRRTEDPEIAOTARPXTKTLNINSIG 115

OY 70 IMKPG-LNALIGPTGGKSLDVLAAKDPGSLGSDVLINGAPRAN-FKNSGVVOD 127

DB 116 ETRDGEIMVILGASGSGKSTLLDALANRAKSLKGTVKLNGTGLQSRMLKVISAYVOD 175

OY 128 DVMGTLTVREMLQPSAALRLATMTNHEKNERINVOELGADKVAASKVGTPIRGVS 187

DB 176 DLFLPMLVTEETLMPFAERFLPSLPSKSKKLVQALIDOLGIRNAKTIIGDEGRGIS 235

OY 188 GGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLKMSKQRTIIFSHOPIY 247

DB 236 GGERRVSIGIDIIHPDILFLDEPTSGDSTAFVWVKVLKRIAQSGSIVMSIHQPSH 295

OY 248 SIFKLPDSTLLASGRIMHGAQELGYFESAGHCEAYNPADFFLDI-----NG 300

DB 296 RVLGDLRLIFLSRGHTVVSGPSALPRFTEFGSPIPENENTFERALDIRELESGAG 355

OY 301 DSTAVLANREEDPKATEIIEPSKODPLIEKLAETIYVNSFYKETKALHQ--LSG 357

DB 356 TRGLIEPNK---KQWEMKQSNRPPLTPP-SSPYNLTIKALIASISRGKLVSGGES 410

```

Qy 358 -----KKKIVFEKISYTTSPFCHOLRWVSKSFKNLGNPQASIAQIIVTVVLGIVGA 411
      |          :          :          :          :          :          :
Db 411 VAHGAGTNTTTLAVAFANPMMEIKITLSKSMNSRRQPELFGIRASVITGTILAT 470
      |          :          :          :          :          :          :
Qy 412 IYFGLKNDSTGIQNRAGVLFELTTCNCFSSVASVELFVVEKKLFIEHYISGYRVSYYL 471
      |          :          :          :          :          :          :
Db 471 VEMRLDNSPKGVQOERIGFPAFAMSTMFYCADLPPVLOERYIFMEETAYNNVRRSSYYL 530
      |          :          :          :          :          :          :
Qy 472 GKLLSDLLPMTLPSIIFTCIYFVMLGPKADAFVMMFTLMVAVSASMAIAAQAQ 531
      |          :          :          :          :          :          :
Db 531 SHAIYS-FPSLLIFLSVAFAATTWAVGLDGLTGLLFYCLIIIASFSWSSSPFTLSGVV 589
      |          :          :          :          :          :          :
Qy 532 SVASVATLMTICFVPMIIFSGILVNLTTIASVLSWLOQFSIPRYGFTALQHNFEFLQNF 591
      |          :          :          :          :          :          :
Db 590 PSMVLCYTIIVAILAVFLFSGFFINRNRIPOVWIMFHMVSLVKRYEAVLQNEFSDARK 649
      |          :          :          :          :          :          :
Qy 592 C--PGINATGNP-----CNYATC--TGEYLVKQG-IDLSP 623
      |          :          :          :          :          :          :
Db 650 CFVRGVQIPDNPPLGELPEVMKLLKLGTVSKSLGVITISSTTCITLTSDDLROGVVQLSK 709
      |          :          :          :          :          :          :
Qy 624 WGLMKHVALACMIVIELTIA---YKLKF 650
      |          :          :          :          :          :          :
Db 710 KN-----CLFIVAFGFFPRILF 727

```

## RESULT 15

```

B96742
Probable ABC transporter F17M19.11 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #ext_change 09-Jul-2004
C:Accession: E96742
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
amnen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaytin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talon,
Ker, M.; Wu, D.; Yu, G.; Frazer, C.W.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; WUID:21016719; PMID:11130712
A:Accession: E96742
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-609 <STO>
A:Cross-references: UNIPROT:Q9C8W6; GB:AE005173; NID:G6978921; PIDN:AAF34313.1; GSPDB:GN
C:Genetics:
A:Gene: F17M19.11
A:Map position: 1
C:Superfamily: Fruit fly white protein, ATP-binding cassette homology

```

```

Query Match          21.3%; Score 712.5; DB 2; Length 609;
Best Local Similarity 31.0%; Pred. No. 4.5e-42;
Matches 190; Conservative 125; Mismatches 256; Indels 41; Gaps 16;

Qy 56 KRPVEKEILSNNGIMKPG-LNALILPTGGKSSLDVLAARKDPSGLSGDVLINGAPRP 114
      |          :          :          :          :          :          :
Db 23 RSTERTILISGYTMSPEFMAVLGPSSGKSTLINAAVAGRLHGSLTGKILINDGKIT 82
      |          :          :          :          :          :          :
Qy 115 ANPKNSGVVQDDVVMGTLTVRENLOFSAALRLATMTNHEKNERINRINVIQELGDKVA 174
      |          :          :          :          :          :          :
Db 83 KQTLKRTGVADDDLLYPLHIVRETLVFAALLRLPRSLTRDYKLAASVISELGLTKCE 142
      |          :          :          :          :          :          :
Qy 175 DSKVGTQPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVALLLKMSK- 233
      |          :          :          :          :          :          :
Db 143 NTVVGNTPIRGISGGERKRTSIAHELLINPSLVLDEPTSGDATAALRLVOTLAGLHG 202
      |          :          :          :          :          :          :
Qy 234 QGRITIIISHQRYSTIFKLPDSLTLASGLMFHGPQDALGYFESAGYHCEAYNNPADF 293
      |          :          :          :          :          :          :
Db 203 KQKTVTSIHOPSSRVFQMFDTVLLSEKCLFVGKRDAMAYFESVGFSPAPPMNPADE 262
      |          :          :          :          :          :          :
Qy 294 FLIDING--DSTAVALNREEDFKATEIIPSPKODKPLIEKLAEL----YVNSFYKETKA 347

```

```

Db 263 ILDLANVCQDTGVTGEREKPNRQTLVTAAYDTLLAPQVTCIEVSHFPQDNARFVKT--- 319
      |          :          :          :          :          :          :
Qy 348 ELHQSLSGEEKKKKIVFEKISYTTSPFCHOL-RWVSKRSFKNLLGNPQASIAQIIVTVVLG 406
      |          :          :          :          :          :          :
Db 320 ---RVNGGGITTCIA-----TWFSQCLILHRLKERRHSEF-----DLRIFOYVAAS 365
      |          :          :          :          :          :          :
Qy 407 LVIGAIYFGLKNDSTGIQNRAGVLFELTTCNCFSSVASVELFVVEKKLFIEHYISGYR 465
      |          :          :          :          :          :          :
Db 366 IICGLMMW--HSDYRDVHDLGLLFFISIFMGVLSFNNAVFTFPQERAFTEBRASGWT 423
      |          :          :          :          :          :          :
Qy 466 VSSYFGLKLLSDLLPMTLPSIIFTCIYFVMLGPKADAFVMMFTLMVAVSASMAIAAQAQ 525
      |          :          :          :          :          :          :
Db 424 LSSYFMAHVLGSLSMELVLPASFLT-FTYVMVYLRPGIIVPFLTLVLLVYLASQGLGL 482
      |          :          :          :          :          :          :
Qy 526 AIAAGOSVSVATLMTICFVPMIIFSGILVNLTTIASVLSWLOQFSIPRYGFTALQHNFEFLQNF 582
      |          :          :          :          :          :          :
Db 483 ALGAAIMPAKKAISTIVYTMALFVLTGGYVN--KVPSSGMVMKVVSTTFYCYRLVAIQ 540
      |          :          :          :          :          :          :
Qy 583 HNEFLGQNFPC--GLNATGNPCNATCTGEEYLVKQIGIDLSPMGLMKHVALACMIVIF 640
      |          :          :          :          :          :          :
Db 541 YGS--GEILRMVLCGDSKKGASAAATSGCRFVEEVI--GDVGMTSVGVLFMLMFRGY 596
      |          :          :          :          :          :          :
Qy 641 LTIAYLKLFLK 652
      |          :          :          :          :          :          :
Db 597 RVLAYLALRRIK 608

```

Search completed: June 6, 2005, 09:44:30  
Job time : 47 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: June 6, 2005, 09:19:13 / Search time 175 Seconds  
(without alignments)  
1916.639 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSSVEVFIPVSGQNTNGF.....MIVFLTAYLKLPLKKYS 655

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: UniProt 03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_crembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3346	99.8	655	2 Q96TA8	Q96TA8 homo sapien
2	3342	99.7	655	2 Q8IX16	Q8IX16 homo sapien
3	3339	99.6	655	1 ABG2_HUMAN	Q9UNQ0 homo sapien
4	3338	99.6	655	2 Q96LDB	Q96LDB homo sapien
5	2849.5	85.0	656	2 Q8MI13	Q8MI13 sus scrofa
6	2762	82.4	657	2 Q7TM55	Q7TM55 mus musculu
7	2761	82.4	541	2 Q86V64	Q86V64 homo sapien
8	2757	82.2	657	2 Q9R004	Q9R004 mus musculu
9	2754	82.2	657	2 Q8OW57	Q8OW57 rattus norv
10	2752	82.1	657	2 Q8OW57	Q8OW57 rattus norv
11	2742	81.8	657	2 Q8OXF3	Q8OXF3 rattus norv
12	1787.5	53.3	650	2 Q8BK15	Q8BK15 mus musculu
13	1786.5	53.3	650	1 ABG3_MOUSE	Q99P81 mus musculu
14	1703.5	50.8	646	2 Q6BH7	Q6BH7 rattus norv
15	864	25.8	1039	2 Q6BIH1	Q6BIH1 debrayomyce
16	863	25.7	645	2 Q6BG61	Q6BG61 parameseliu
17	847.5	25.3	801	2 Q8T691	Q8T691 dictyosteli
18	835.5	24.9	1049	1 ADP1_YEAST	P25371 saccharomyc
19	826.5	24.7	1011	2 Q756R4	Q756R4 ashyba goss
20	821.5	24.5	1078	2 Q7SHS0	Q7SHS0 neurospora
21	819	24.4	1055	2 Q6FKY0	Q6FKY0 candida gla
22	812	24.2	687	1 WHIT_DROME	P10090 drosophila
23	809	24.1	751	2 Q9JY54	Q9JY54 arabidopsi
24	808	24.1	687	2 Q9NH94	Q9NH94 bombyx mori
25	806	24.0	687	2 Q94960	Q94960 drosophila
26	802.5	23.9	1022	2 Q6CS24	Q6CS24 kluyveromyc
27	800.5	23.9	737	2 Q9FT51	Q9FT51 arabidopsi
28	793	23.7	679	2 Q9BH97	Q9BH97 ceratilis c
29	790	23.6	692	2 P91892	P91892 aedes aegy
30	783.5	23.4	648	2 Q9C6M5	Q9C6M5 arabidopsi
31	777.5	23.2	679	2 Q8IS30	Q8IS30 bacterocera

32	776.5	23.2	567	2 Q9EG17	Q9EG17 arabidopsi
33	774	23.1	646	2 Q9C6R7	Q9C6R7 arabidopsi
34	772	23.0	670	2 Q77423	Q77423 bacterocera
35	772	23.0	679	1 WHIT_CERCA	Q17320 ceratilis c
36	767.5	22.9	687	2 Q9C8R2	Q9C8R2 arabidopsi
37	766.5	22.9	695	2 Q8IMQ5	Q8IMQ5 oryza sativ
38	763.5	22.8	798	2 Q8T689	Q8T689 dictyosteli
39	759	22.6	703	2 Q8RXN0	Q8RXN0 arabidopsi
40	755	22.5	672	2 Q9LI82	Q9LI82 arabidopsi
41	754.5	22.5	677	1 WHIT_LUCCU	Q05360 lucilia cup
42	751	22.4	669	2 Q8WRP2	Q8WRP2 tribolium c
43	751	22.4	695	1 WHIT_ANOGA	Q27256 anopheles g
44	749.5	22.4	691	2 Q8RW19	Q8RW19 arabidopsi
45	746	22.3	649	2 Q9SIT6	Q9SIT6 arabidopsi

## ALIGNMENTS

RESULT 1  
Q96TA8 PRELIMINARY; PRT; 655 AA.  
AC Q96TA8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE ATP-binding cassette superfamily G (white) member 2 (ATP-binding cassette, sub-family G, member 2).  
GN Name=ABCG2;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21201983; PubMed=11306452;  
RA Komatani H., Kotani H., Hara Y., Nakagawa R., Matsumoto M.,  
RA Arakawa H., Nishimura S.;  
RT Identification of breast cancer resistant protein/mitoxantone  
RT resistance/placenta-specific, ATP-binding cassette transporter as a  
RT transporter of NB-506 and J-107088, topoisomerase I inhibitors with an  
RT indolocarbazole structure.";  
RT Cancer Res. 61:2827-2832(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Pancreeas;  
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnaa.242603899;  
RA Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altchul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Martins K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinici P., Scheetz T.B.,  
RA Rabb S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalski U., Smalios D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Pancreeas;  
RC Strussberg R.;  
RA Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.  
RL -1- SIMILARITY: Belongs to the ABC transporter family.

DR EMBL; AB051855; BAB46933.1; -.  
DR EMBL; BC021281; AAH21281.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR006162; Pnntne\_S.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.  
KW ATP-binding.  
SQ SEQUENCE 655 AA; 72313 MW; ABAF6B96034C3A8 CRC64;  
  
Query Match 99.8%; Score 3346; DB 2; Length 655;  
Best Local Similarity 99.8%; Pred. No. 1.2e-208;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYKLSGFLPCRKYV 60  
DB 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYKLSGFLPCRKYV 60  
QY 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANPKCN 120  
DB 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANPKCN 120  
QY 121 SGVYVDDVVMGTLTYRENLOFSALRLATTTMNEKNERINRVIOELGLDKVADSKVGT 180  
DB 121 SGVYVDDVVMGTLTYRENLOFSALRLATTTMNEKNERINRVIOELGLDKVADSKVGT 180  
QY 181 QPFRGVSGGERKRTSIMGELITDPSILFLDEPTTGDSSTANAVLLLRKMSKQGTIIF 240  
DB 181 QPFRGVSGGERKRTSIMGELITDPSILFLDEPTTGDSSTANAVLLLRKMSKQGTIIF 240  
QY 241 SIHOPRYSIFKLPDSITLLASGRLMFHGPAQELAGYFESAGYHCEAYNNPADFFLIING 300  
DB 241 SIHOPRYSIFKLPDSITLLASGRLMFHGPAQELAGYFESAGYHCEAYNNPADFFLIING 300  
QY 301 DSTAVALNREDEDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360  
DB 301 DSTAVALNREDEDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360  
QY 361 ITYFKKISTTSCHQLRWKRSFKNLGNPQASIAQIIVTVVLGLVIGAIYFGKND 420  
DB 361 ITYFKKISTTSCHQLRWKRSFKNLGNPQASIAQIIVTVVLGLVIGAIYFGKND 420  
QY 421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVVEKKLFHEHISGYRVSSTYFLGLSDLLP 480  
DB 421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVVEKKLFHEHISGYRVSSTYFLGLSDLLP 480  
QY 481 MTMLPSIIFTCTIYFNLGLKPKADAFVMMFTLMVAAYASSMALAAAGOSVVSATLL 540  
DB 481 MTMLPSIIFTCTIYFNLGLKPKADAFVMMFTLMVAAYASSMALAAAGOSVVSATLL 540  
QY 541 MTICFVFMNIFSGLLVNLTTIASWLSWLOYSIPRIGFTALQNEHFLGONFCGLNATGN 600  
DB 541 MTICFVFMNIFSGLLVNLTTIASWLSWLOYSIPRIGFTALQNEHFLGONFCGLNATGN 600  
QY 601 NPCNVAATCGEEVLYVQGITDLSFPGIMKONVALACMIVIFLTAIYKTLFLKKYS 655  
DB 601 NPCNVAATCGEEVLYVQGITDLSFPGIMKONVALACMIVIFLTAIYKTLFLKKYS 655  
  
RESULT 2  
Q81X16 PRELIMINARY; PRT; 655 AA.  
Q81X16; 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE ATP-binding cassette protein ABCG2.  
GN Name=ABCG2;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
RX [1]  
RP SEQUENCE FROM N.A.  
RA Yoshikawa M., Yabuchi H., Ikegami Y., Ishikawa T.;  
RU Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
CC -i- SIMILARITY: Belongs to the ABC transporter family.  
DR EMBL; AF463519; AA014617.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR006162; Pnntne\_S.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.  
KW ATP-binding.  
SQ SEQUENCE 655 AA; 72314 MW; ABAF6B591D4C5A8 CRC64;  
  
Query Match 99.7%; Score 3342; DB 2; Length 655;  
Best Local Similarity 99.7%; Pred. No. 2.2e-208;  
Matches 653; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYKLSGFLPCRKYV 60  
DB 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICRYKLSGFLPCRKYV 60  
QY 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANPKCN 120  
DB 61 KEILSININGIMKPGNALIIGPTGGKSSLDVLAARKDPSGLSGDVLINGAPRANPKCN 120  
QY 121 SGVYVDDVVMGTLTYRENLOFSALRLATTTMNEKNERINRVIOELGLDKVADSKVGT 180  
DB 121 SGVYVDDVVMGTLTYRENLOFSALRLATTTMNEKNERINRVIOELGLDKVADSKVGT 180  
QY 181 QPFRGVSGGERKRTSIMGELITDPSILFLDEPTTGDSSTANAVLLLRKMSKQGTIIF 240  
DB 181 QPFRGVSGGERKRTSIMGELITDPSILFLDEPTTGDSSTANAVLLLRKMSKQGTIIF 240  
QY 241 SIHOPRYSIFKLPDSITLLASGRLMFHGPAQELAGYFESAGYHCEAYNNPADFFLIING 300  
DB 241 SIHOPRYSIFKLPDSITLLASGRLMFHGPAQELAGYFESAGYHCEAYNNPADFFLIING 300  
QY 301 DSTAVALNREDEDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360  
DB 301 DSTAVALNREDEDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGEKKK 360  
QY 361 ITYFKKISTTSCHQLRWKRSFKNLGNPQASIAQIIVTVVLGLVIGAIYFGKND 420  
DB 361 ITYFKKISTTSCHQLRWKRSFKNLGNPQASIAQIIVTVVLGLVIGAIYFGKND 420  
QY 421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVVEKKLFHEHISGYRVSSTYFLGLSDLLP 480  
DB 421 TGIQNRAGVLFLLTNNQCFSSVSAVELFVVEKKLFHEHISGYRVSSTYFLGLSDLLP 480  
QY 481 MTMLPSIIFTCTIYFNLGLKPKADAFVMMFTLMVAAYASSMALAAAGOSVVSATLL 540  
DB 481 MTMLPSIIFTCTIYFNLGLKPKADAFVMMFTLMVAAYASSMALAAAGOSVVSATLL 540  
QY 541 MTICFVFMNIFSGLLVNLTTIASWLSWLOYSIPRIGFTALQNEHFLGONFCGLNATGN 600  
DB 541 MTICFVFMNIFSGLLVNLTTIASWLSWLOYSIPRIGFTALQNEHFLGONFCGLNATGN 600



OY 601 NECNATCTGEEYLVKQIDLSPMGLMKNNHVALACMIVFLTAYLKLFKYS 655  
 DB 601 NECNATCTGEEYLVKQIDLSPMGLMKNNHVALACMIVFLTAYLKLFKYS 655  
 RESULT 3  
 ID ABG2 HUMAN STANDARD; PRT; 655 AA.  
 AC 09UN00; 095374; 09BY73; 09NUS0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).  
 GN Name:ABCG2; Synonyms:ABCG, BCRP, BCRP1;  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RC TISSUE=Placenta;  
 RX MEDLINE=99065313; PubMed=9850061;  
 RA Alikmetes R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;  
 RT "A human placenta-specific ATP-binding cassette gene (ABCG) on chromosome 4q32 that is involved in multidrug resistance.";  
 RL Cancer Res. 56:5337-5339(1998).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RX TISSUE=Breast cancer;  
 RA MEDLINE=99080071; PubMed=9861027; DOI=10.1073/pnas.95.26.15665;  
 RT Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K., Rose D.D.;  
 RL "A multidrug resistance transporter from human MCF-7 breast cancer cells.";  
 RN [3]  
 RC Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).  
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K., Rose D.D.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).  
 RN [4]  
 RC SEQUENCE FROM N.A.  
 RA Kage K., Tsukuhara S., Sugiyama T., Asada S., Ishikawa E., Teurro T., Sugimoto Y.;  
 RL "Breast cancer resistance protein constitutes a 140-kDa complex as a homodimer.";  
 RN [5]  
 RC Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RC SEQUENCE OF 198-655 FROM N.A.  
 RX TISSUE=Placenta; DOI=10.1038/ng1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakematsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Oabayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S., Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahata K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A., Sudo H., Hosoiri T., Kaku T., Kodaira H., Kondo H., Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Furuya T., Ohta Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Ohtsuka Y., Tanai H., Kimata M., Watanabe S., Yoshida M., Hattori T., Ieshida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hattori T., Kueno J., Kanehori T., Takahashi-Fujii A., Hara R., Tanase T.-O., Nomura Y., Togiyama S., Komai F., Hara R., Takeuchi K., Arita M., Imose N., Musashino K., Yuki H., Oshita A., Sasaki N., Aotsuka S., Yoshihara Y., Matsumura H., Ichihara T., Shitahara N., Sano S., Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O., Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H., Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., Ohtsuka R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M., RA Togashi T., Oyama M., Hata H., Watanabe M., Konatsu T., Nakagawa K., Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K., Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita K., Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 RN [6]  
 RX REVIEW.  
 RA MEDLINE=21474438; PubMed=11590207;  
 RT Schmitz G., Langmann T., Heimerl S.;  
 RL "Role of ABCG and other ABCG family members in lipid metabolism.";  
 J. Lipid Res. 42:1513-1520(2001).  
 RN [7]  
 RA VARIANTS LEU-431 AND LEU-489.  
 RA Itoda M., Saito Y., Shira K., Minami H., Ohtsu A., Yoshida T., Saito N., Suzuki H., Sugiyama Y., Ozawa S., Sawada J.-I.;  
 RT "Eight novel single nucleotide polymorphisms in ABCG2/BCRP in Japanese cancer patients administered irinotecan.";  
 RL Drug Metab. Pharmacokinet. 18:212-217(2003).  
 CC -1- FUNCTION: Xenobiotic transporter that appears to play a major role in the multidrug resistance phenotype of a specific MCF-7 breast cancer cell line. When overexpressed, the transfected cells become resistant to mitoxantrone, daunorubicin and doxorubicin, display diminished intracellular accumulation of daunorubicin, and manifest an ATP-dependent increase in the efflux of rhodamine 123.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -1- SIMILARITY: Belongs to the ABC transporter family. ABCG (White) subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL, AF103796; AAD09188.1; -  
 DR EMBL, AF098951; AAC97367.1; -  
 DR EMBL, AB056867; BAB39212.1; -  
 DR EMBL, AK002040; BAA92050.1; -  
 DR Genew; HGNC:74; ABCG2.  
 DR MIM; 603756; -  
 DR GO; GO:0016021; C:integral to membrane; TAS.  
 DR GO; GO:0005524; F:ATP binding; TAS.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . . ; TAS.  
 DR GO; GO:0005215; F:transporter activity; TAS.  
 DR GO; GO:0008559; F:xenobiotic-transporting ATPase activity; TAS.  
 DR GO; GO:0042493; F:response to drug; TAS.  
 DR GO; GO:0006810; P:transport; TAS.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_Transporter.  
 DR Pfam; PF00005; ABC\_tran\_1.  
 DR ProDom; PD00006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; FALSE\_NEG.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Polymorphism; Transmembrane; Transport.  
 FT DOMAIN 1 395  
 FT TRANSMEM 396 416  
 FT TRANSMEM 417 428  
 FT TRANSMEM 429 448  
 FT TRANSMEM 450 477  
 FT TRANSMEM 478 498  
 FT TRANSMEM 499 506  
 FT TRANSMEM 507 527  
 FT TRANSMEM 528 535  
 FT TRANSMEM 536 556  
 FT DOMAIN 557 630  
 FT TRANSMEM 557 630  
 FT TRANSMEM 630 655  
 FT TRANSMEM 656 677  
 FT TRANSMEM 678 699  
 FT TRANSMEM 700 721  
 FT TRANSMEM 722 743  
 FT TRANSMEM 744 765  
 FT TRANSMEM 766 787  
 FT TRANSMEM 788 809  
 FT TRANSMEM 810 831  
 FT TRANSMEM 832 853  
 FT TRANSMEM 854 875  
 FT TRANSMEM 876 897  
 FT TRANSMEM 898 919  
 FT TRANSMEM 920 941  
 FT TRANSMEM 942 963  
 FT TRANSMEM 964 985  
 FT TRANSMEM 986 1007  
 FT TRANSMEM 1008 1029  
 FT TRANSMEM 1030 1051  
 FT TRANSMEM 1052 1073  
 FT TRANSMEM 1074 1095  
 FT TRANSMEM 1096 1117  
 FT TRANSMEM 1118 1139  
 FT TRANSMEM 1140 1161  
 FT TRANSMEM 1162 1183  
 FT TRANSMEM 1184 1205  
 FT TRANSMEM 1206 1227  
 FT TRANSMEM 1228 1249  
 FT TRANSMEM 1250 1271  
 FT TRANSMEM 1272 1293  
 FT TRANSMEM 1294 1315  
 FT TRANSMEM 1316 1337  
 FT TRANSMEM 1338 1359  
 FT TRANSMEM 1360 1381  
 FT TRANSMEM 1382 1403  
 FT TRANSMEM 1404 1425  
 FT TRANSMEM 1426 1447  
 FT TRANSMEM 1448 1469  
 FT TRANSMEM 1470 1491  
 FT TRANSMEM 1492 1513  
 FT TRANSMEM 1514 1535  
 FT TRANSMEM 1536 1557  
 FT TRANSMEM 1558 1579  
 FT TRANSMEM 1580 1601  
 FT TRANSMEM 1602 1623  
 FT TRANSMEM 1624 1645  
 FT TRANSMEM 1646 1667  
 FT TRANSMEM 1668 1689  
 FT TRANSMEM 1690 1711  
 FT TRANSMEM 1712 1733  
 FT TRANSMEM 1734 1755  
 FT TRANSMEM 1756 1777  
 FT TRANSMEM 1778 1799  
 FT TRANSMEM 1800 1821  
 FT TRANSMEM 1822 1843  
 FT TRANSMEM 1844 1865  
 FT TRANSMEM 1866 1887  
 FT TRANSMEM 1888 1909  
 FT TRANSMEM 1910 1931  
 FT TRANSMEM 1932 1953  
 FT TRANSMEM 1954 1975  
 FT TRANSMEM 1976 1997  
 FT TRANSMEM 1998 2019  
 FT TRANSMEM 2020 2041  
 FT TRANSMEM 2042 2063  
 FT TRANSMEM 2064 2085  
 FT TRANSMEM 2086 2107  
 FT TRANSMEM 2108 2129  
 FT TRANSMEM 2130 2151  
 FT TRANSMEM 2152 2173  
 FT TRANSMEM 2174 2195  
 FT TRANSMEM 2196 2217  
 FT TRANSMEM 2218 2239  
 FT TRANSMEM 2240 2261  
 FT TRANSMEM 2262 2283  
 FT TRANSMEM 2284 2305  
 FT TRANSMEM 2306 2327  
 FT TRANSMEM 2328 2349  
 FT TRANSMEM 2350 2371  
 FT TRANSMEM 2372 2393  
 FT TRANSMEM 2394 2415  
 FT TRANSMEM 2416 2437  
 FT TRANSMEM 2438 2459  
 FT TRANSMEM 2460 2481  
 FT TRANSMEM 2482 2503  
 FT TRANSMEM 2504 2525  
 FT TRANSMEM 2526 2547  
 FT TRANSMEM 2548 2569  
 FT TRANSMEM 2570 2591  
 FT TRANSMEM 2592 2613  
 FT TRANSMEM 2614 2635  
 FT TRANSMEM 2636 2657  
 FT TRANSMEM 2658 2679  
 FT TRANSMEM 2680 2701  
 FT TRANSMEM 2702 2723  
 FT TRANSMEM 2724 2745  
 FT TRANSMEM 2746 2767  
 FT TRANSMEM 2768 2789  
 FT TRANSMEM 2790 2811  
 FT TRANSMEM 2812 2833  
 FT TRANSMEM 2834 2855  
 FT TRANSMEM 2856 2877  
 FT TRANSMEM 2878 2899  
 FT TRANSMEM 2900 2921  
 FT TRANSMEM 2922 2943  
 FT TRANSMEM 2944 2965  
 FT TRANSMEM 2966 2987  
 FT TRANSMEM 2988 3009  
 FT TRANSMEM 3010 3031  
 FT TRANSMEM 3032 3053  
 FT TRANSMEM 3054 3075  
 FT TRANSMEM 3076 3097  
 FT TRANSMEM 3098 3119  
 FT TRANSMEM 3120 3141  
 FT TRANSMEM 3142 3163  
 FT TRANSMEM 3164 3185  
 FT TRANSMEM 3186 3207  
 FT TRANSMEM 3208 3229  
 FT TRANSMEM 3230 3251  
 FT TRANSMEM 3252 3273  
 FT TRANSMEM 3274 3295  
 FT TRANSMEM 3296 3317  
 FT TRANSMEM 3318 3339  
 FT TRANSMEM 3340 3361  
 FT TRANSMEM 3362 3383  
 FT TRANSMEM 3384 3405  
 FT TRANSMEM 3406 3427  
 FT TRANSMEM 3428 3449  
 FT TRANSMEM 3450 3471  
 FT TRANSMEM 3472 3493  
 FT TRANSMEM 3494 3515  
 FT TRANSMEM 3516 3537  
 FT TRANSMEM 3538 3559  
 FT TRANSMEM 3560 3581  
 FT TRANSMEM 3582 3603  
 FT TRANSMEM 3604 3625  
 FT TRANSMEM 3626 3647  
 FT TRANSMEM 3648 3669  
 FT TRANSMEM 3670 3691  
 FT TRANSMEM 3692 3713  
 FT TRANSMEM 3714 3735  
 FT TRANSMEM 3736 3757  
 FT TRANSMEM 3758 3779  
 FT TRANSMEM 3780 3801  
 FT TRANSMEM 3802 3823  
 FT TRANSMEM 3824 3845  
 FT TRANSMEM 3846 3867  
 FT TRANSMEM 3868 3889  
 FT TRANSMEM 3890 3911  
 FT TRANSMEM 3912 3933  
 FT TRANSMEM 3934 3955  
 FT TRANSMEM 3956 3977  
 FT TRANSMEM 3978 3999  
 FT TRANSMEM 4000 4021  
 FT TRANSMEM 4022 4043  
 FT TRANSMEM 4044 4065  
 FT TRANSMEM 4066 4087  
 FT TRANSMEM 4088 4109  
 FT TRANSMEM 4110 4131  
 FT TRANSMEM 4132 4153  
 FT TRANSMEM 4154 4175  
 FT TRANSMEM 4176 4197  
 FT TRANSMEM 4198 4219  
 FT TRANSMEM 4220 4241  
 FT TRANSMEM 4242 4263  
 FT TRANSMEM 4264 4285  
 FT TRANSMEM 4286 4307  
 FT TRANSMEM 4308 4329  
 FT TRANSMEM 4330 4351  
 FT TRANSMEM 4352 4373  
 FT TRANSMEM 4374 4395  
 FT TRANSMEM 4396 4416  
 FT TRANSMEM 4417 4438  
 FT TRANSMEM 4439 4460  
 FT TRANSMEM 4461 4482  
 FT TRANSMEM 4483 4504  
 FT TRANSMEM 4505 4525  
 FT TRANSMEM 4526 4547  
 FT TRANSMEM 4548 4569  
 FT TRANSMEM 4570 4591  
 FT TRANSMEM 4592 4612  
 FT TRANSMEM 4613 4634  
 FT TRANSMEM 4635 4655  
 FT TRANSMEM 4656 4677  
 FT TRANSMEM 4678 4698  
 FT TRANSMEM 4699 4720  
 FT TRANSMEM 4721 4742  
 FT TRANSMEM 4743 4764  
 FT TRANSMEM 4765 4785  
 FT TRANSMEM 4786 4807  
 FT TRANSMEM 4808 4829  
 FT TRANSMEM 4830 4851  
 FT TRANSMEM 4852 4872  
 FT TRANSMEM 4873 4894  
 FT TRANSMEM 4895 4915  
 FT TRANSMEM 4916 4937  
 FT TRANSMEM 4938 4959  
 FT TRANSMEM 4960 4981  
 FT TRANSMEM 4982 5003  
 FT TRANSMEM 5004 5024  
 FT TRANSMEM 5025 5046  
 FT TRANSMEM 5047 5068  
 FT TRANSMEM 5069 5089  
 FT TRANSMEM 5090 5110  
 FT TRANSMEM 5111 5131  
 FT TRANSMEM 5132 5152  
 FT TRANSMEM 5153 5173  
 FT TRANSMEM 5174 5194  
 FT TRANSMEM 5195 5215  
 FT TRANSMEM 5216 5236  
 FT TRANSMEM 5237 5257  
 FT TRANSMEM 5258 5278  
 FT TRANSMEM 5279 5299  
 FT TRANSMEM 5300 5321  
 FT TRANSMEM 5322 5342  
 FT TRANSMEM 5343 5363  
 FT TRANSMEM 5364 5385  
 FT TRANSMEM 5386 5406  
 FT TRANSMEM 5407 5427  
 FT TRANSMEM 5428 5448  
 FT TRANSMEM 5449 5470  
 FT TRANSMEM 5471 5491  
 FT TRANSMEM 5492 5512  
 FT TRANSMEM 5513 5533  
 FT TRANSMEM 5534 5554  
 FT TRANSMEM 5555 5575  
 FT TRANSMEM 5576 5596  
 FT TRANSMEM 5597 5617  
 FT TRANSMEM 5618 5638  
 FT TRANSMEM 5639 5660  
 FT TRANSMEM 5661 5681  
 FT TRANSMEM 5682 5702  
 FT TRANSMEM 5703 5723  
 FT TRANSMEM 5724 5744  
 FT TRANSMEM 5745 5765  
 FT TRANSMEM 5766 5786  
 FT TRANSMEM 5787 5807  
 FT TRANSMEM 5808 5828  
 FT TRANSMEM 5829 5849  
 FT TRANSMEM 5850 5870  
 FT TRANSMEM 5871 5891  
 FT TRANSMEM 5892 5912  
 FT TRANSMEM 5913 5933  
 FT TRANSMEM 5934 5954  
 FT TRANSMEM 5955 5975  
 FT TRANSMEM 5976 5996  
 FT TRANSMEM 5997 6017  
 FT TRANSMEM 6018 6038  
 FT TRANSMEM 6039 6059  
 FT TRANSMEM 6060 6080  
 FT TRANSMEM 6081 6101  
 FT TRANSMEM 6102 6122  
 FT TRANSMEM 6123 6143  
 FT TRANSMEM 6144 6164  
 FT TRANSMEM 6165 6185  
 FT TRANSMEM 6186 6206  
 FT TRANSMEM 6207 6227  
 FT TRANSMEM 6228 6248  
 FT TRANSMEM 6249 6269  
 FT TRANSMEM 6270 6290  
 FT TRANSMEM 6291 6311  
 FT TRANSMEM 6312 6332  
 FT TRANSMEM 6333 6353  
 FT TRANSMEM 6354 6374  
 FT TRANSMEM 6375 6395  
 FT TRANSMEM 6396 6416  
 FT TRANSMEM 6417 6437  
 FT TRANSMEM 6438 6458  
 FT TRANSMEM 6459 6479  
 FT TRANSMEM 6480 6500  
 FT TRANSMEM 6501 6521  
 FT TRANSMEM 6522 6542  
 FT TRANSMEM 6543 6563  
 FT TRANSMEM 6564 6584  
 FT TRANSMEM 6585 6605  
 FT TRANSMEM 6606 6626  
 FT TRANSMEM 6627 6647  
 FT TRANSMEM 6648 6668  
 FT TRANSMEM 6669 6689  
 FT TRANSMEM 6690 6710  
 FT TRANSMEM 6711 6730  
 FT TRANSMEM 6731 6751  
 FT TRANSMEM 6752 6772  
 FT TRANSMEM 6773 6793  
 FT TRANSMEM 6794 6814  
 FT TRANSMEM 6815 6835  
 FT TRANSMEM 6836 6856  
 FT TRANSMEM 6857 6877  
 FT TRANSMEM 6878 6898  
 FT TRANSMEM 6899 6919  
 FT TRANSMEM 6920 6940  
 FT TRANSMEM 6941 6961  
 FT TRANSMEM 6962 6982  
 FT TRANSMEM 6983 7003  
 FT TRANSMEM 7004 7024  
 FT TRANSMEM 7025 7045  
 FT TRANSMEM 7046 7066  
 FT TRANSMEM 7067 7087  
 FT TRANSMEM 7088 7108  
 FT TRANSMEM 7109 7129  
 FT TRANSMEM 7130 7150  
 FT TRANSMEM 7151 7171  
 FT TRANSMEM 7172 7192  
 FT TRANSMEM 7193 7213  
 FT TRANSMEM 7214 7234  
 FT TRANSMEM 7235 7255  
 FT TRANSMEM 7256 7276  
 FT TRANSMEM 7277 7297  
 FT TRANSMEM 7298 7318  
 FT TRANSMEM 7319 7339  
 FT TRANSMEM 7340 7360  
 FT TRANSMEM 7361 7381  
 FT TRANSMEM 7382 7402  
 FT TRANSMEM 7403 7423  
 FT TRANSMEM 7424 7444  
 FT TRANSMEM 7445 7465  
 FT TRANSMEM 7466 7486  
 FT TRANSMEM 7487 7507  
 FT TRANSMEM 7508 7528  
 FT TRANSMEM 7529 7549  
 FT TRANSMEM 7550 7570  
 FT TRANSMEM 7571 7591  
 FT TRANSMEM 7592 7612  
 FT TRANSMEM 7613 7633  
 FT TRANSMEM 7634 7654  
 FT TRANSMEM 7655 7675  
 FT TRANSMEM 7676 7696  
 FT TRANSMEM 7697 7717  
 FT TRANSMEM 7718 7738  
 FT TRANSMEM 7739 7759  
 FT TRANSMEM 7760 7780  
 FT TRANSMEM 7781 7801  
 FT TRANSMEM 7802 7822  
 FT TRANSMEM 7823 7843  
 FT TRANSMEM 7844 7864  
 FT TRANSMEM 7865 7885  
 FT TRANSMEM 7886 7906  
 FT TRANSMEM 7907 7927  
 FT TRANSMEM 7928 7948  
 FT TRANSMEM 7949 7969  
 FT TRANSMEM 7970 7990  
 FT TRANSMEM 7991 8011  
 FT TRANSMEM 8012 8032  
 FT TRANSMEM 8033 8053  
 FT TRANSMEM 8054 8074  
 FT TRANSMEM 8075 8095  
 FT TRANSMEM 8096 8116  
 FT TRANSMEM 8117 8137  
 FT TRANSMEM 8138 8158  
 FT TRANSMEM 8159 8179  
 FT TRANSMEM 8180 8200  
 FT TRANSMEM 8201 8221  
 FT TRANSMEM 8222 8242  
 FT TRANSMEM 8243 8263  
 FT TRANSMEM 8264 8284  
 FT TRANSMEM 8285 8305  
 FT TRANSMEM 8306 8326  
 FT TRANSMEM 8327 8347  
 FT TRANSMEM 8348 8368  
 FT TRANSMEM 8369 8389  
 FT TRANSMEM 8390 8410  
 FT TRANSMEM 8411 8431  
 FT TRANSMEM 8432 8452  
 FT TRANSMEM 8453 8473  
 FT TRANSMEM 8474 8494  
 FT TRANSMEM 8495 8515  
 FT TRANSMEM 8516 8536  
 FT TRANSMEM 8537 8557  
 FT TRANSMEM 8558 8578  
 FT TRANSMEM 8579 8599  
 FT TRANSMEM 8600 8621  
 FT TRANSMEM 8622 8642  
 FT TRANSMEM 8643 8663  
 FT TRANSMEM 8664 8684  
 FT TRANSMEM 8685 8705  
 FT TRANSMEM 8706 8726  
 FT TRANSMEM 8727 8747  
 FT TRANSMEM 8748 8768  
 FT TRANSMEM 8769 8789  
 FT TRANSMEM 8790 8810  
 FT TRANSMEM 8811 8831  
 FT TRANSMEM 8832 8852  
 FT TRANSMEM 8853 8873  
 FT TRANSMEM 8874 8894  
 FT TRANSMEM 8895 8915  
 FT TRANSMEM 8916 8936  
 FT TRANSMEM 8937 8957  
 FT TRANSMEM 8958 8978  
 FT TRANSMEM 8979 8999  
 FT TRANSMEM 9000 9020  
 FT TRANSMEM 9021 9041  
 FT TRANSMEM 9042 9062  
 FT TRANSMEM 9063 9083  
 FT TRANSMEM 9084 9104  
 FT TRANSMEM 9105 9125  
 FT TRANSMEM 9126 9146  
 FT TRANSMEM 9147 9167  
 FT TRANSMEM 9168 9188  
 FT TRANSMEM 9189 9209  
 FT TRANSMEM 9210 9230  
 FT TRANSMEM 9231 9251  
 FT TRANSMEM 9252 9272  
 FT TRANSMEM 9273 9293  
 FT TRANSMEM 9294 9314  
 FT TRANSMEM 9315 9335  
 FT TRANSMEM 9336 9356  
 FT TRANSMEM 9357 9377  
 FT TRANSMEM 9378 9398  
 FT TRANSMEM 9399 9419  
 FT TRANSMEM 9420 9440  
 FT TRANSMEM 9441 9461  
 FT TRANSMEM 9462 9482  
 FT TRANSMEM 9483 9503  
 FT TRANSMEM 9504 9524  
 FT TRANSMEM 9525 9545  
 FT TRANSMEM 9546 9566  
 FT TRANSMEM 9567 9587  
 FT TRANSMEM 9588 9608  
 FT TRANSMEM 9609 9629  
 FT TRANSMEM 9630 9650  
 FT TRANSMEM 9651 9671  
 FT TRANSMEM 9672 9692  
 FT TRANSMEM 9693 9713  
 FT TRANSMEM 9714 9734  
 FT TRANSMEM 9735 9755  
 FT TRANSMEM 9756 9776  
 FT TRANSMEM 9777 9797  
 FT TRANSMEM 9798 9818  
 FT TRANSMEM 9819 9839  
 FT TRANSMEM 9840 9860  
 FT TRANSMEM 9861 9881  
 FT TRANSMEM 9882 9902  
 FT TRANSMEM 9903 9923  
 FT TRANSMEM 9924 9944  
 FT TRANSMEM 9945 9965  
 FT TRANSMEM 9966 9986  
 FT TRANSMEM 9987 10000

```
FT TRANSMEM 631 651 Potential.
FT DOMAIN 652 655 Cytoplasmic (Potential) .
FT NP_BIND 80 87 ATP (Potential) .
FT CARBOHYD 418 N-linked (GLNAC. . .) (Potential) .
FT CARBOHYD 557 557 N-linked (GLNAC. . .) (Potential) .
FT CARBOHYD 556 596 N-linked (GLNAC. . .) (Potential) .
FT VARIANT 431 431 F -> L.
FT VARIANT 431 431 /Frid=VAR_018349.
FT VARIANT 489 489 F -> L.
FT VARIANT 489 489 /Frid=VAR_018350.
FT CONFLICT 24 24 V -> A (in Ref. 2 and 4) .
FT CONFLICT 166 166 E -> Q (in Ref. 2 and 4) .
FT CONFLICT 208 208 F -> S (in Ref. 1) .
FT CONFLICT 315 316 Missing (in Ref. 5) .
FT CONFLICT 482 482 R -> T (in Ref. 2) .
SQ SEQUENCE 655 AA; 72343 MW; 89A6D3511DC5CCE0 CRC64;

Query Match 99.6%; Score 3339; DB 1; Length 655;
Best Local Similarity 99.5%; Pred. No. 3,4e-208;
Matches 653; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSSSNEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICYRVKLSGFLPCRKPYE 60
DB 1 MSSSNEVFIPVSGQNTNGFPATVSNLKAFTGAVLSFHNICYRVKLSGFLPCRKPYE 60
QY 61 KEILSNINGIMKRGNAIIGPTGGKSSLLDVLAARKDPGSLGSDVLINGAPRPANPKCN 120
DB 61 KEILSNINGIMKRGNAIIGPTGGKSSLLDVLAARKDPGSLGSDVLINGAPRPANPKCN 120
QY 121 SGVYVDDVVMGTLVRENLOFSAALRLATMTNHEKNERINRVIOELGLDYADSKVGT 180
DB 121 SGVYVDDVVMGTLVRENLOFSAALRLATMTNHEKNERINRVIOELGLDYADSKVGT 180
QY 181 QFIRGVSGGRRKSTSGMELITDPSILFLDEPTTGLDSSSTANAVLLKMKMSQGRITIF 240
DB 181 QFIRGVSGGRRKSTSGMELITDPSILFLDEPTTGLDSSSTANAVLLKMKMSQGRITIF 240
QY 241 SIHQPRYSIFPKLDSITLTLASGRLMFHGPAQOALGYFESAGYHCEAVNNPADPFLDING 300
DB 241 SIHQPRYSIFPKLDSITLTLASGRLMFHGPAQOALGYFESAGYHCEAVNNPADPFLDING 300
QY 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAEHLQLSGGEKKK 360
DB 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAEHLQLSGGEKKK 360
QY 361 ITVFKESITSTSCCHOLRWYSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGKNDK 420
DB 361 ITVFKESITSTSCCHOLRWYSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGKNDK 420
QY 421 TGIQNRAGVLFPLTNQCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDLPL 480
DB 421 TGIQNRAGVLFPLTNQCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDLPL 480
QY 481 MTMLPSIIFTCIYFPMGLKPKADAFVMMFTLMMVAYSASSMALAIAGOSVVSATLL 540
DB 481 MTMLPSIIFTCIYFPMGLKPKADAFVMMFTLMMVAYSASSMALAIAGOSVVSATLL 540
QY 541 MTICPFPMIIFSGILVNLTTIASWLSWLOFYFSIPRYGFTALQNHFEIGQNFCEGLNATGN 600
DB 541 MTICPFPMIIFSGILVNLTTIASWLSWLOFYFSIPRYGFTALQNHFEIGQNFCEGLNATGN 600
QY 601 NPCNVATCTGEAYLVKQIDLSPMGLMKHNAVALACMIVIFLTTIAYLKLFLKKYS 655
DB 601 NPCNVATCTGEAYLVKQIDLSPMGLMKHNAVALACMIVIFLTTIAYLKLFLKKYS 655

RESULT 4
Q96LD6 PRELIMINARY; PRT; 655 AA.
AC Q96LD6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
```

```
DE ABC transporter ABCG2.
GN Name=ABCG2;
OS Homo sapiens (human) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21424790; PubMed=11533706;
RA Zhou S., Schuetz J.D., Bunting K.D., Colapietro A.M., Sampath J.,
RA Morris J.J., Lagutina I., Grosveld G.C., Osawa M., Nakauchi H.,
RA Sorrentino B.P.;
RT "The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of
RT stem cells and is a molecular determinant of the side-population
RT phenotype."
RT Nat. Med. 7:1028-1034(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Schuetz J.D., Wall A.M., Sampath J., Sorrentino B., Du G.;
RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL: AY017168; AAC52982.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO: GO:0000166; F:nucleotide binding; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR006162; Ppatlne_S.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00893; ABC_TRANSPORTER 2; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR KW ATP-binding.
SQ SEQUENCE 655 AA; 72287 MW; B3B5DC020C95C4A8 CRC64;

Query Match 99.6%; Score 3338; DB 2; Length 655;
Best Local Similarity 99.7%; Pred. No. 4e-208;
Matches 653; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSSSNEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICYRVKLSGFLPCRKPYE 60
DB 1 MSSSNEVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICYRVKLSGFLPCRKPYE 60
QY 61 KEILSNINGIMKRGNAIIGPTGGKSSLLDVLAARKDPGSLGSDVLINGAPRPANPKCN 120
DB 61 KEILSNINGIMKRGNAIIGPTGGKSSLLDVLAARKDPGSLGSDVLINGAPRPANPKCN 120
QY 121 SGVYVDDVVMGTLVRENLOFSAALRLATMTNHEKNERINRVIOELGLDYADSKVGT 180
DB 121 SGVYVDDVVMGTLVRENLOFSAALRLATMTNHEKNERINRVIOELGLDYADSKVGT 180
QY 181 QFIRGVSGGRRKSTSGMELITDPSILFLDEPTTGLDSSSTANAVLLKMKMSQGRITIF 240
DB 181 QFIRGVSGGRRKSTSGMELITDPSILFLDEPTTGLDSSSTANAVLLKMKMSQGRITIF 240
QY 241 SIHQPRYSIFPKLDSITLTLASGRLMFHGPAQOALGYFESAGYHCEAVNNPADPFLDING 300
DB 241 SIHQPRYSIFPKLDSITLTLASGRLMFHGPAQOALGYFESAGYHCEAVNNPADPFLDING 300
QY 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAEHLQLSGGEKKK 360
DB 301 DSTAVALNREDEPKATEIIEPSKODKPLIEKLAIEIVNSSFYKETAEHLQLSGGEKKK 360
QY 361 ITVFKESITSTSCCHOLRWYSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGKNDK 420
DB 361 ITVFKESITSTSCCHOLRWYSKRSFKNLGNPOASIAQIIVTVVLGLVIGAIYFGKNDK 420
QY 421 TGIQNRAGVLFPLTNQCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDLPL 480
DB 421 TGIQNRAGVLFPLTNQCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDLPL 480
```

Oy		481	MMLSLIIPTCTVYFNLGKPKRADAFVMMFTLMMVAVSASSALALIAAGQSVASVATLL	540
Dd		461	NRMLSIITCTCYFMFLGAKADAFVMMFTLMMVAVSASSALALIAAGQSVASVATLL	540
Oy		541	MTICVFPMWIFSGLLVNLTTLTASWLSMWLOYSPIPRGFALQHNEFLGONFCPLGNATGN	600
Dd		541	MTICVFPMWIFSGLLVNLTTLTASWLSMWLOYSPIPRGFALQHNEFLGONFCPLGNATGN	600
Oy		601	NPENATCTGESELYLKQIGIDLSPWGLMKHVALACIVIFLTAYIKLIPLKKYS	655
Dd		601	NPENATCTGESELYLKQIGIDLSPWGLMKHVALACIVIFLTAYIKLIPLKKYS	655
	RESULT 5			
	QBMB3			
ID	QBMB3	PRELIMINARY;	PRT:	656 AA.
AC	QBMB3;			
DT	01-OCT-2002 (TREMBLrel. 22, Created)			
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)			
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
DE	Brain multidrug resistance protein.			
GN	Name=BMDP;			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.			
NCBI	NCBI_TaxId=96823;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=22050127; PubMed=12054514; DOI=10.1016/S0006-291X(02)00376-5;			
RA	Eisenblatter T., Galla H.J.;			
RT	"A new multidrug resistance protein at the blood-brain barrier."			
RL	Biochem. Biophys. Res. Commun. 293:1273-1278(2002)."			
CC	-1- SIMILARITY: Belongs to the ABC transporter family.			
DR	EMBL; AJ420927; CAD12785.1; -.			
DR	PIR; JC7860; JC7860.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0042628; F:ATPase activity, coupled to transmembrane m. . .; IEA.			
DR	GO; GO:0000168; F:nucleotide binding; IEA.			
DR	InterPro; IPR003593; AAA_ATPase.			
DR	InterPro; IPR003439; ABC_transporter.			
DR	InterPro; IPR006162; Pntantne_S.			
DR	Pfam; PF000005; ABC_tran; 1.			
DR	ProDom; PD000006; ABC_transporter; 1.			
DR	SMART; SM00382; AAA; 1.			
DR	PROSITE; PS00893; ABC_TRANSPORTER_2; 1.			
DR	PROSITE; PS00012; PHOSPHOPANTHEINE; UNKNOWN_1.			
KV	ATP-binding.			
SO	SEQUENCE	656 AA;	72391 MW;	118ADSB53D9067 CRG64;
	Query Match	85.0%;	Score 2849.5;	DB 2; Length 656;
	Best Local Similarity	84.3%;	Pred. No. 2e-176;	
	Matches 553;	Conservative 44;	Mismatches 58;	Indels 1; Gaps 1
Oy	1	MSSNVVEVIPISQCNTPFPATSNNDLCAPFGAVLSPFNHCYRKLSGFLPCCKPYE	60	
Dd	1	MSSNSYQVISPSKNKTNGLPSSSHNELTTSAGVALSHDHDCYRVKVASGFLPCRKYIE	60	
Oy	61	KEILSINGIMKPGNALIIGPTGGGKSLLDVLAARKDPBGLSDVLINGARRPANFKCN	120	
Dd	61	KEILTININGIMKPGNALIIGPTGGGKSLLDVLAARKDPBGLSDVLINGARRPANFKCN	120	
Oy	121	SGYVVQDDVMGTLLVRENLFQSALARLEPTTMTHNEKMRIMVIOELGDKVADSKVGT	180	
Dd	121	SGYVVQDDVMGTLLVRENLFQSALARLEPTTMTHNEKMRIMVIOELGDKVADSKVGT	180	
Oy	181	QPIRGVSGGERKRSTISGMELTDPSILFDPEPTGDSSTANAVALLLRMSKGRTIIF	240	
Dd	181	QPIRGVSGGERKRSTISAMELITDPSILFDPEPTGDSSTANAVALLLRMSKGRTIIF	240	
Oy	241	SIHQPRYSIFKLDELTLIASGRMFHGPAQELGYFESAGYCEAYNNPADFLDIING	300	

Db	STHPRYS1F.KL.FD.SU.LTLAS.R.L.F.HG.FARE.L.GY.FASI.GINCE.PYNN.PAD.F.FLDVING	3000
Db	241	
Qy	301	
Db	301	
Qy	360	
Db	361	
Qy	420	
Db	421	
Qy	480	
Db	481	
Qy	540	
Db	541	
Qy	600	
Db	601	
RESULT 6		
Q7TMS5	PRELIMINARY;	PRT; 657 AA.
ID	Q7TMS5	
AC	01-OCT-2003 (Tremblrel. 25, Created)	
DT	01-OCT-2003 (Tremblrel. 25, Last sequence update)	
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)	
DE	ATP-binding cassette, sub-family G, member 2.	
GN	Name=Adcg2;	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxId=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;	
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;	
RA	Strasbourg R.L., Feingold E.A., Grouse L.H., Derge J.G.	
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schler G.D.,	
RA	Alteich L.F., Zeeberg B., Buecaw K.H., Schaefer C.F., Bhat N.K.,	
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,	
RA	Diatchenko L., Marushina K., Farmer A.M., Rubin G.M., Hong L.,	
RA	Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scherz T.E.,	
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Plange C.,	
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullen S.J.,	
RA	Boeck S.A., McEwan P.J., McKernan K.D., Malek J.A., Gunaratne P.H.,	
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,	
RA	Villalon D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,	
RA	Faney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,	
RA	Whiting M., Madan A., Young A.C., Shechenko Y., Bouffard G.G.,	
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,	
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,	
RA	Krzywnicki M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,	
RA	Jones S.J., Maria M.A.,	
RT	"Generation and initial analysis of more than 15,000 full-length human	
RT	and mouse cDNA sequences."	
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;	
RA	Strasbourg R.;	
RA	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.	
CC	-1- SIMILARITY: Belongs to the ABC transporter family.	
DR	EMBL, BC053730.1; --	
GO	GO:0016021; C:integral to membrane; TAS.	

DR InterPro: IPR003593; AAA ATPase.  
DR InterPro: IPR003439; ABC transporter.  
DR InterPro: IPR006162; Pntane\_S.  
DR Pfam: PF00005; ABC\_tran; 1.  
DR ProDom: PD000006; ABC\_transporter; 1.  
DR SMART: SM00382; AAA; 1.  
DR PROSITE: PS00893; ABC\_TRANSPORTER\_2; 1.  
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.  
KM ATP-binding.  
SQ SEQUENCE 657 AA; 72977 MW; DCDY0CSD9FA2BA5F CRC64;

Query Match 82.4%; Score 2762; DB 2; Length 657;  
Best Local Similarity 81.5%; Pred. No. 9,4e-171;  
Matches 536; Conservative 52; Mismatches 66; Indels 4; Gaps 3;

QY 1 MSSNNVEFIPVQSGNTNGPPTASNDLKAFTEGAVLSFINICRYVKLSGFLPCRPVE 60  
DB 1 MSSNNHVLVPMQSRNNNGLPRTNSRAVRLTAGDVLSPFHITRYRVKSGFL-VRKTV 59  
QY 61 KEILSINGIMKPGNALIGPTGGKSSLDVLAARDPGLSGDVLINGAPPPANPKCN 120  
DB 60 KEILSDINGIMKGLNAILGPTGGKSSLDVLAARDPGLSGDVLINGAPPPANPKCC 119  
QY 121 SGVYVDDVVMGTLVRENLOFSAALRLATMTNHEKNERINRVIOELGLDKVADSKVGT 180  
DB 120 SGVYVDDVVMGTLVRENLOFSAALRLPTMKHEKNERINTIKELGLEKVAADSKVGT 179  
QY 181 QPFRGYSGGERKRTSGIMEITDPSILFDEPTTGLDSTANNAVLLLLKMSKQGRITIF 240  
DB 180 QPFRGISGGERKRTSGIMEITDPSILFDEPTTGLDSTANNAVLLLLKMSKQGRITIF 239  
QY 241 SIHQPRYSIFPKLPDSLTLLASGRLMFHPAOEALGYESAGYCEAVNNPADFLDING 300  
DB 240 SIHQPRYSIFPKLPDSLTLLASGRLVFRGPAOKALEYASGYCEPNNPADFLDING 299  
QY 301 DSTAVLANREE-DPKATELIEPSKODKPLIEKLAETIVNSSFYKETVAELHQLSGEKKK 359  
DB 300 DSSAVMLNREQNEANKTEPEKRPVENISEFYINGAIYGETVAELDQLPGAQEEK 359  
QY 360 KIVPFEISITTSFCHQLRWVSKSPKLNLPQASIAQIIVTVVLGLVIGALYFGKND 419  
DB 360 GTSAFKEPYVVISFCHQLRWIRARSPKLNLPQASVAQIIVTVVLGLVIGALYFDLKYD 419  
QY 420 STGIQRAGVLFPLTNQCSSSAVELFVVEKKLFIEHYISGYRVSSYFLGLKLSDDL 479  
DB 420 AAGMQRAGVLFPLTNQCSSSAVELFVVEKKLFIEHYISGYRVSSYFLGRKMSDDL 479  
QY 480 PMTMLPSIIFTCIYVFMGLKPKADAFVVMFTLMVAVYASSSMALAIAGQSVSVATL 539  
DB 480 PMFLPSVIFTCVLYFMGLKKTVDAFIMFTLIMVAYASSSMALAIAGQSVSVATL 539  
QY 540 LMTICVFPMIFSGGLVNLTTIASWSLQYFSIPRGFTALQHNELGONFCGLNATG 599  
DB 540 LMTIAVFPFMMIFSGGLVNLTTIGPWSLQYFSIPRGFTALQNEFLGQFCPGFVNTD 599  
QY 600 NNPC-NYATCTGEEYLVKQIDLSPWGLMKNHVALACMIVIFLTIAVLLFLPKKY 655  
DB 600 NSTCNSYALCTGNEYLINGISLSPWGLMKNHVALACMIIIFLTIAVLLFLPKKY 657

RESULT 7  
Q86V64 PRELIMINARY; PRT; 541 AA.  
AC Q86V64;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-MAR-2003 (TrEMBLrel. 24, Last sequence update)  
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE ATP-binding cassette sub-family G member 2 (Fragment).  
GN Name=ABCG2;  
OS Homo sapiens (human).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
NCBI\_TaxId=9606;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22959505; Pubmed=12958161;  
RA Zhang W., Mojilovic-Petrovic J., Andrade M.F., Zhang H., Ball M.,  
RT Stanimirovic D.B.;  
RT "The expression and functional characterization of ABCG2 in brain  
endothelial cells and vessels";  
RL FASEB J. 17:2085-2087(2003).  
CC -1- SIMILARITY: Belongs to the ABC transporter family.  
DR EMBL: AY288307; AAP31310.1; -  
DR GO: GO:0016020; C:membrane; IEA.  
DR GO: GO:0005524; F:ATP binding; IEA.  
DR GO: GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.  
DR GO: GO:0006810; P:transport; IEA.  
DR InterPro: IPR003439; ABC\_transporter.  
DR InterPro: IPR006162; Pntane\_S.  
DR Pfam: PF00005; ABC\_tran; 1.  
DR ProDom: PD000006; ABC\_transporter; 1.  
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.  
KM ATP-binding.  
FT NON TER 1 1  
SQ SEQUENCE 541 AA; 60450 MW; 050B2742C6A3F66C CRC64;

Query Match 82.4%; Score 2761; DB 2; Length 541;  
Best Local Similarity 99.8%; Pred. No. 8,6e-171;  
Matches 540; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 115 ANPKNSGVYVDDVVMGTLVRENLOFSAALRLATMTNHEKNERINRVIOELGLDKVA 174  
DB 1 ANPKNSGVYVDDVVMGTLVRENLOFSAALRLATMTNHEKNERINRVIOELGLDKVA 60  
QY 175 DSKVQFPFRGYSGGERKRTSGIMEITDPSILFDEPTTGLDSTANNAVLLLLKMSKQ 234  
DB 61 DSKVQFPFRGYSGGERKRTSGIMEITDPSILFDEPTTGLDSTANNAVLLLLKMSKQ 120  
QY 235 GRTIIFSIHQPRYSIFPKLPDSLTLLASGRLMFHPAOEALGYESAGYCEAVNNPADFF 294  
DB 121 GRTIIFSIHQPRYSIFPKLPDSLTLLASGRLMFHPAOEALGYESAGYCEAVNNPADFF 180  
QY 295 LDIINDSTAVLANREDEKATEIIEPSKODKPLIEKLAETIVNSSFYKETVAELHQLSG 354  
DB 181 LDIINDSTAVLANREDEKATEIIEPSKODKPLIEKLAETIVNSSFYKETVAELHQLSG 240  
QY 355 GEKKKKTITFEKESITTSFCHQLRWVSKSPKLNLPQASIAQIIVTVVLGLVIGALYF 414  
DB 241 GEKKKKTITFEKESITTSFCHQLRWVSKSPKLNLPQASIAQIIVTVVLGLVIGALYF 300  
QY 415 GLKNDSTGIQNRAGVLFPLTNQCSSSAVELFVVEKKLFIEHYISGYRVSSYFLGLK 474  
DB 301 GLKNDSTGIQNRAGVLFPLTNQCSSSAVELFVVEKKLFIEHYISGYRVSSYFLGLK 360  
QY 475 LSDLLPMTMLPSIIFTCIYVFMGLKPKADAFVVMFTLMVAVYASSSMALAIAGQSVY 534  
DB 361 LSDLLPMTMLPSIIFTCIYVFMGLKPKADAFVVMFTLMVAVYASSSMALAIAGQSVY 420  
QY 535 SVATLMTICVFPMIFSGGLVNLTTIASWSLQYFSIPRGFTALQHNELGONFCPG 594  
DB 421 SVATLMTICVFPMIFSGGLVNLTTIASWSLQYFSIPRGFTALQHNELGONFCPG 480  
QY 595 LNAATGNPCNYATCTGEEYLVKQIDLSPWGLMKNHVALACMIVIFLTIAVLLFLPKKY 654  
DB 481 LNAATGNPCNYATCTGEEYLVKQIDLSPWGLMKNHVALACMIVIFLTIAVLLFLPKKY 540

RESULT 8  
Q9R004 PRELIMINARY; PRT; 657 AA.  
ID Q9R004;  
AC Q9R004;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Breast cancer resistance protein 1.  
 GN Name=Abcg2; Synonyms=Bcrpl;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB; TISSUE=Liver;  
 RA MEDLINE=99413474; PubMed=10485464;  
 RX Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.;  
 RT "The mouse Bcrpl/Mxr/Abcg gene: amplification and overexpression in  
 cell lines selected for resistance to topotecan, mitoxantrone, or  
 doxorubicin."  
 RL Cancer Res. 59:4237-4241(1999).  
 CC -1- SIMILARITY: Belongs to the ABC transporter family.  
 DR EMBL; AF140218; AAC54216.1; -.  
 DR MGD; MGI:1347061; Abcg2.  
 DR GO; GO:0016021; C:integral to membrane; TAS.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006162; Pntane\_S.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN\_1.  
 DR ATP-binding.  
 KW SEQUENCE 657 AA; 73021 MW; 207B70BC272CC0D5 CRC64;  
 SQ

Query Match 82.2%; Score 2757; DB 2; Length 657;  
 Best Local Similarity, 81.5%; Pred. No. 2e-170; Indels 4; Gaps 3;  
 Matches 536; Conservative 51; Mismatches 67;

QY 1 MSSNVEEPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRKLSGFLPCKRKYE 60  
 DB 1 MSSNDHVLVPMSSQNNKGLPMSRAVTTAEGDVLSTHHTYKVKSGFL-VRKTYE 59

QY 61 KEILSNGIMKPGINALILGPTGGKSSLLDYLARKDPSGSLVINGARPNANFKCN 120  
 DB 60 KEILSDINGIMKPGINALILGPTGGKSSLLDYLARKDPSGSLVINGARPNANFKCC 119

QY 121 SGYVVDVDMGTLVRENLOFSALRLATMTNHEKNERINRVIOELGLDVAASKYGT 180  
 DB 120 SGYVVDVDMGTLVRENLOFSALRLPTTMKNEKNERINTTIIKEIGLEKVAASKYGT 179

QY 181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKRMKSGKRTIIF 240  
 DB 180 QPIRGISGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKRMKSGKRTIIF 239

QY 241 SIHQPRYSIFKLFDLSLTLLASGRLMFHGPAQALGFSSAGYHCEAYNNPADFLDIING 300  
 DB 240 SIHQPRYSIFKLFDLSLTLLASGRLMFHGPAQALGFSSAGYHCEAYNNPADFLDIING 299

QY 301 DSTAVALLNR-DFKATEIIEPSKODKPLIEKLAETIYNSSFYKTKAEHLQSGEKKK 359  
 DB 300 DSSAVMLNRBEDONANKTEEPSKGEKPYIEMISEFYINSALYGETKELDLPQAQEKK 359

QY 360 KITVEKEISYTTSPFCHOLRWKSPKSLGNPQASIQIITVVLGLVIGVIGKLNKD 419  
 DB 360 GTSARKEPYVTSFCHOLRWKSPKSLGNPQASIQIITVVLGLVIGVIGVIGKLNKD 419

QY 420 STGIQNRAGVLEFLTTNQCFSVSAVELFVVEKELFIEHYISGYRVSSYFLGKLLSDL 479  
 DB 420 AAGQNRAGVLEFLTTNQCFSVSAVELFVVEKELFIEHYISGYRVSSYFLGKLLSDL 479

QY 480 PWTMLPSIITFCIVFPMGLKPKADAFPMFTLMMVAVSASSMALALIAAGSVSVATL 539  
 DB 480 PWRFLPSVITFCILFPMGLKKTVAFFIMFTLLMVAVTASSMALALATGSSVSVATL 539

QY 540 LMTICFVFMIMISGLLVNLTIASMLSWLQYFSIDRYGFTALQHNBFIGNPCGINTG 599

DB 540 LMTICFVFMIMISGLLVNLTIASMLSWLQYFSIDRYGFTALQYNBFIGNPCGINTG 599

QY 600 NNPC--NATCGEEYLVKQGIKLSPKGLMKVHALACIVFLITAYLKLFLKYS 655  
 DB 600 NSTCVNSYALCTGNEYLLNIGIELSPWGLMKVHALACIITIFLITAYLKLFLKYS 657

RESULT 9  
 ID Q80W57 PRELIMINARY; PRT; 657 AA.  
 AC Q80W57;  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE ABC transporter ABCG2.  
 GN Name=Abcg2;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mistar;  
 RA Hori S., Ohtsuki S., Tachikawa M., Kimura N., Kondo T., Watanabe M.,  
 RA Nakashima E., Terasaki T.;  
 RT "Functional expression of rat ABCG2 on the luminal side of brain  
 capillaries and its enhancement by astrocyte-derived soluble  
 factor(s)."  
 RL J. Neurochem. 90:526-536(2004).  
 CC -1- SIMILARITY: Belongs to the ABC transporter family.  
 DR EMBL; AB105817; BAC76396.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006162; Pntane\_S.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN\_1.  
 DR ATP-binding.  
 KW SEQUENCE 657 AA; 72960 MW; C975C61A08489027 CRC64;  
 SQ

Query Match 82.2%; Score 2754; DB 2; Length 657;  
 Best Local Similarity, 81.0%; Pred. No. 3.1e-170; Indels 4; Gaps 3;  
 Matches 533; Conservative 52; Mismatches 69;

QY 1 MSSNVEEPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRKLSGFLPCKRKYE 60  
 DB 1 MSSNDHVLVPMSSQNNKGLPMSRGARTLAEGDVLSTHHTYKVKSGFL-VRKTYE 59

QY 61 KEILSNGIMKPGINALILGPTGGKSSLLDYLARKDPSGSLVINGARPNANFKCN 120  
 DB 60 KEILSDINGIMKPGINALILGPTGGKSSLLDYLARKDPSGSLVINGARPNANFKCS 119

QY 121 SGYVVDVDMGTLVRENLOFSALRLATMTNHEKNERINRVIOELGLDVAASKYGT 180  
 DB 120 SGYVVDVDMGTLVRENLOFSALRLPKMKTHEKNERINTTIIKEIGLEKVAASKYGT 179

QY 181 QPIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKRMKSGKRTIIF 240  
 DB 180 QPIRGISGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKRMKSGKRTIIF 239

QY 241 SIHQPRYSIFKLFDLSLTLLASGRLMFHGPAQALGFSSAGYHCEAYNNPADFLDIING 300  
 DB 240 SIHQPRYSIFKLFDLSLTLLASGRLMFHGPAQALGFSSAGYHCEAYNNPADFLDIING 299

QY 301 DSTAVALLNR-BEDFKATEIIEPSKODKPLIEKLAETIYNSSFYKTKAEHLQSGEKKK 359

```
Db 300 DSSAVMLNRGEQHANKTEBPSREKPIIENLAEPFINSITTYGETAELDQLPVAKKK 359
Qy 360 KITVFEISYTTSPCHQLRWVSRKSPFNLLGNPOASIAQIIVTVVLGVLGAIYFGKND 419
Db 360 GSSAFREPVVTSFCHQLRWIARSFPKNLLGNPOASIAQIIVTVVLGVLGAIYFGKND 419
Qy 420 STGIQNRAGVLPFLTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 479
Db 420 PTGMQNRAGVFPFLTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 479
Qy 480 PMTMLPSIIFTCIYVFMGLGKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATL 539
Db 480 PMRFLPSVIYTCILLYFMGLGKRTVEAFVMMFTLMVAYSASSMALAIAAGQSVSVATL 539
Qy 540 LMTICFVFMVIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQPCGLNATG 599
Db 540 LMTISFVFMVIFSGLLVNLTTIGWLSWLOYSIPRYGFTALQHNFLGQPCGLNATG 599
Qy 600 NNPC--NYATCTGEEVLVKGIDLSPWGLMKHNVALACMIVLTITAYLKLFLKYS 655
Db 600 NSTCVNSYITCTGNDYILNQGIDLSPWGLMRNVHVALACMIIIFLTITAYLKLFLKYS 657
```

## RESULT 10

Q80ST1 PRELIMINARY; PRT; 657 AA.

```
AC Q80ST1;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE ATP-binding cassette protein G2 transcript variant B (ATP-binding
DE cassette protein G2 transcript variant C) (ATP-binding cassette
DE protein G2 transcript variant A).
GN Name=Abcg2;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RA Yabuchi H., Ishikawa T.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL: AY089996; AAM09106.1; -.
DR EMBL: AY089997; AAM09107.1; -.
DR EMBL: AY089998; AAM09108.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO: GO:0000166; F:nucleotide binding; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR006162; Pntane_S.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PSS0893; ABC_TRANSPORTER_2; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR ATP-binding.
KW
SQ SEQUENCE 657 AA; 72960 MW; E194871E1C1AC201 CRC64;
```

Query Match 82.1%; Score 2752; DB 2; Length 657;

Best Local Similarity 80.9%; Pred. No. 4.2e-170;

Matches 532; Conservative 53; Mismatches 69; Indels 4; Gaps 3;

```
Qy 1 MSSSNVEFIPVQSGNTNGFPATASNDLKAFTEGAVLSFNNICRYVLLKSGFLPCRPRVE 60
Db 1 MSSSNHVLVPMQGRNKNGIPGKSSRGARLAEQDVLSFHITRYVYKSGFL-VKRTAE 59
Qy 61 KEILSINGIMKGLNALIIGPTGGKSSLLDVLAAKDPGSLGQDVLINGAPRPANFKCN 120
```

```
Db 60 KEILSINGIMKGLNALIIGPTGGKSSLLDVLAAKDPGSLGQDVLINGAPRPANFKCS 119
Qy 121 SGYVODDVMGLTYRENLOFSALRLATMTNHEKNERINRVIQELGDKYADSVGT 180
Db 120 SGYVODDVMGLTYRENLOFSALRLPRAMTKHEKNERINTIIEELGDKYADSVGT 179
Qy 181 QFTRVSGGERKRTSIGMEIITDPSILFDEPTTGLDSSTANAVLLLLKMSKQRTIIF 240
Db 180 QFTRVSGGERKRTSIGMEIITDPSILFDEPTTGLDSSTANAVLLLLKMSKQRTIIF 239
Qy 241 SIHQPRYSIFKLFDSTLLASGLMFRGQELGYPESGHCYCEAYNNPADFFLIING 300
Db 240 SIHQPRYSIFKLFDSTLLASGLMFRGQELGYPESGHCYCEAYNNPADFFLIING 299
Qy 301 DSTAVVLMNR-EEDFKATFETIEPSKODKPLIEKAEIYVNSSPFKETKAEHLQSGEKK 359
Db 300 DSSAVMLNRGEQHANKTEBPSREKPIIENLAEPFINSITTYGETAELDQLPVAKKK 359
Qy 360 KITVFEISYTTSPCHQLRWVSRKSPFNLLGNPOASIAQIIVTVVLGVLGAIYFGKND 419
Db 360 GSSAFREPVVTSFCHQLRWIARSFPKNLLGNPOASIAQIIVTVVLGVLGAIYFGKND 419
Qy 420 STGIQNRAGVLPFLTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 479
Db 420 PTGMQNRAGVFPFLTNOCFSSVSAVELFVVEKKLFIEHYISGYRVSSYFLGKLSDDL 479
Qy 480 PMTMLPSIIFTCIYVFMGLGKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATL 539
Db 480 PMRFLPSVIYTCILLYFMGLGKRTVEAFVMMFTLMVAYSASSMALAIAAGQSVSVATL 539
Qy 540 LMTICFVFMVIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGQPCGLNATG 599
Db 540 LMTISFVFMVIFSGLLVNLRTIGWLSWLOYSIPRYGFTALQHNFLGQPCGLNATG 599
Qy 600 NNPC--NYATCTGEEVLVKGIDLSPWGLMKHNVALACMIVLTITAYLKLFLKYS 655
Db 600 NSTCVNSYITCTGNDYILNQGIDLSPWGLMRNVHVALACMIIIFLTITAYLKLFLKYS 657
```

## RESULT 11

Q80XF3 PRELIMINARY; PRT; 657 AA.

```
AC Q80XF3;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE ATP-binding cassette transporter ABCG2.
GN Name=Abcg2;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Shitomo K., Satake M., Okaya A., Kitahara J., Kitahara N.,
RA Takemura M., Sakagami M., Terada N., Tsujimura T.;
RL "Hepatic Oval Cells Have the Side Population Phenotype Defined by
RT Expression of ATP-binding Cassette Transporter ABCG2/BCRP1.";
RL Am. J. Pathol. 0:0-0(2003).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL: AB094089; BAC75666.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO: GO:0000166; F:nucleotide binding; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR006162; Pntane_S.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
```





Db 1 MASNNDPTVISMIERHLCDEPTNTSDKLTTEBAVLSFHNIISYQETVSGFPLRKKAYV 60  
 Qy 61 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAAKOPSGLSGVDLNGAPRANPKCN 120  
 Db 61 IERLSNISGIMKPGALNAILGPTGGKSSLLDVLAAKOPSGLSGVDLNGAPRANPKCN 120  
 Qy 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRYIOELGDKVADSKVGT 180  
 Db 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRYIOELGDKVADSKVGT 180  
 Qy 181 OFIRGVSGGRKRTSICMELITDPSILFLDEPTTGLDSSANAVLLLLKXMSKQRTIIF 240  
 Db 174 QNIRKPKSKELRKRSTIMBELVTEHPILFLDDPTTGLRTTDTTLVLRKMSKGRITIF 233  
 Qy 241 SIHOPRYSIFKLPDSLTLLASGRLMFHPAQOALGYESAGYHCEAVNPNADFLDIING 300  
 Db 234 SINOPQYSIFKLPDSLTLLASGRLMFHPAQOALGYESAGYHCEAVNPNADFLDIING 293  
 Qy 301 DSTAVALNRE--EDFKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGEK 357  
 Db 294 GFSNIIIDTEBDGHEDDKYEELEFEROY--VTGKLAMNVAQSPLYSETRAIIIDQLGEOK 350  
 Qy 358 KKKITVKEISYTTSPFCHOLRWVSKSPKRLGNPQASIAQIIVTVLGLVIGAIYGLK 417  
 Db 351 LERSASV-ETTCVTPFCHOLRWVSKSPKRLGNPQASIAQIIVTVLGLVIGAIYGLK 409  
 Qy 418 NDSTGIONRAGVLFELTTCNOCSSVSAVELFVVEKLFHEIYISGYRVSSYFGLKLS 477  
 Db 410 NDCEVQMRAGLYLTLTFCITISVSAGELFVIDRVPFLHEHTSGYRVSSYFGLKLS 469  
 Qy 478 LLEPTMLPSIIFTCIYFVFMGLKPKADAFVMMFTLMMVVASASMAIAAAGSVVSA 537  
 Db 470 LIRPLRLPSTVFSILITVIVAGVMSKCFFTMCTIIVLAVSASPLISIGAGENAVAVP 529  
 Qy 538 TLLMTICFVFMFISGLLVNLTIAWSLMOYFSIRYGFALQHNELFQNCFCGLNA 597  
 Db 530 TLLMTICFVFMFISGLLVNLTIAWSLMOYFSIRYGFALQHNELFQNCFCGLNA 589  
 Qy 598 TGNNPC-NVATCTGEEYLVKQIGIDSPMGMLKNHVALACMIVFLTAAYLKLFLK 652  
 Db 590 EYVSRCHNVIYCTGEFLMIOGIDLSMGEMENHVALVCMIILITTYVQLQOVK 645

RESULT 13  
 ABG3 MOUSE STANDARD; PRT; 650 AA.  
 AC 099P81;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 3.  
 GN Name=Abcg3;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;  
 RX MEDLINE=21030753; Pubmed=11178751;  
 RA Mckley L., Jain P., Miyake K., Schriml L.M., Rao K., Fojo T.,  
 RA Bates S., Dean M.;  
 RT "An ATP-binding cassette gene (ABCG3) closely related to the multidrug  
 transporter ABCG2 (MXR/ABCP) has an unusual ATP-binding domain.";  
 RL Mamm. Genome 12:86-88(2001).  
 CC -1- SUBUNIT: May dimerize with another subunit to form a functional  
 CC transporter.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -1- TISSUE SPECIFICITY: Highest levels of expression in thymus and  
 CC spleen. Detected in lung and small intestine.  
 CC -1- SIMILARITY: Belongs to the ABC transporter family. ABCG (white)  
 CC subfamily.  
 CC -1- CAUTION: Seems to have a defective ATP-binding region.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF324242; AKL14241.1; -.  
 DR MGD; MGI:1351624; Abcg3.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; FALSE\_NEG.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Transmembrane; Transport.  
 FT DOMAIN 1 387 Cytoplasmic (Potential).  
 FT TRANSMEM 388 408 1 (Potential).  
 FT DOMAIN 409 420 Extracellular (Potential).  
 FT TRANSMEM 421 441 2 (Potential).  
 FT DOMAIN 442 469 Cytoplasmic (Potential).  
 FT TRANSMEM 470 490 3 (Potential).  
 FT DOMAIN 491 498 Extracellular (Potential).  
 FT TRANSMEM 499 519 4 (Potential).  
 FT DOMAIN 520 527 Cytoplasmic (Potential).  
 FT TRANSMEM 528 548 5 (Potential).  
 FT DOMAIN 549 623 Extracellular (Potential).  
 FT TRANSMEM 624 644 6 (Potential).  
 FT DOMAIN 645 648 Cytoplasmic (Potential).  
 SQ SEQUENCE 650 AA; 73623 MW; 86A5AB4DD26971C CRC64;  
 Query Match 53.3%; Score 1786.5; DB 1; Length 650;  
 Best local Similarity 55.3%; Pred. No. 1.9e-107;  
 Matches 363; Conservative 104; Mismatches 174; Indels 15; Gaps 5;  
 Qy 1 MSSNNEVPIPVSGNTNGPATAINDLKAFTGAVLSFHNIISYQETVSGFPLRKKAYV 60  
 Db 1 MASNNDPTVISMIERHLCDEPTNTSDKLTTEBAVLSFHNIISYQETVSGFPLRKKAYV 60  
 Qy 61 KEILSNINGIMKPGALNAILGPTGGKSSLLDVLAAKOPSGLSGVDLNGAPRANPKCN 120  
 Db 61 IERLSNISGIMKPGALNAILGPTGGKSSLLDVLAAKOPSGLSGVDLNGAPRANPKCN 120  
 Qy 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRYIOELGDKVADSKVGT 180  
 Db 121 SGYVQDDVVMGTLTRENLOFSAALRLATMTNHEKNERINRYIOELGDKVADSKVGT 180  
 Qy 181 OFIRGVSGGRKRTSICMELITDPSILFLDEPTTGLDSSANAVLLLLKXMSKQRTIIF 240  
 Db 174 QNIRKPKSKELRKRSTIMBELVTEHPILFLDDPTTGLRTTDTTLVLRKMSKGRITIF 233  
 Qy 241 SIHOPRYSIFKLPDSLTLLASGRLMFHPAQOALGYESAGYHCEAVNPNADFLDIING 300  
 Db 234 SINOPQYSIFKLPDSLTLLASGRLMFHPAQOALGYESAGYHCEAVNPNADFLDIING 293  
 Qy 301 DSTAVALNRE--EDFKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGEK 357  
 Db 294 GFSNIIIDTEBDGHEDDKYEELEFEROY--VTGKLAMNVAQSPLYSETRAIIIDQLGEOK 350  
 Qy 358 KKKITVKEISYTTSPFCHOLRWVSKSPKRLGNPQASIAQIIVTVLGLVIGAIYGLK 417  
 Db 351 LERSASV-ETTCVTPFCHOLRWVSKSPKRLGNPQASIAQIIVTVLGLVIGAIYGLK 409  
 Qy 418 NDSTGIONRAGVLFELTTCNOCSSVSAVELFVVEKLFHEIYISGYRVSSYFGLKLS 477  
 Db 410 NDCEVQMRAGLYLTLTFCITISVSAGELFVIDRVPFLHEHTSGYRVSSYFGLKLS 469  
 Qy 478 LLEPTMLPSIIFTCIYFVFMGLKPKADAFVMMFTLMMVVASASMAIAAAGSVVSA 537  
 Db 470 LIRPLRLPSTVFSILITVIVAGVMSKCFFTMCTIIVLAVSASPLISIGAGENAVAVP 529  
 Qy 538 TLLMTICFVFMFISGLLVNLTIAWSLMOYFSIRYGFALQHNELFQNCFCGLNA 597



Db 530 TLVTVTVFVFMFLFSGSLSYLPGSFLPKJSMWQYFSLPHYGFPALLHNEFLGONFCBEHNT 589  
 Qy 558 TGNMNC-NVATCTGEBYLVKQIDISPMGLMKHVALACMIYVFLTIAYLKLLFLK 652  
 Db 590 EVSRCHNVVICTGEBFLMIQIGIDISLWSGFWENHVALVCTMIITLITVOLLQVK 645

## RESULT 14

Q68HW7 PRELIMINARY; PRT; 646 AA.  
 ID Q68HW7  
 AC Q68HW7;  
 DT 25-OCT-2004 (TREMBlrel. 28, Created)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE ABC transporter ABCG3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RA Zhang W., Bissoc-Hagani S., Yam Y., Stanimirovic D.B.;  
 RT "Cloning and characterization of ABCG3 in rat brain endothelial  
 cells."  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF688113; AAT99308.1; -  
 DR InterPro; IPR003439; ABC transporter.  
 DR Pfam; PF00005; ABC tran. 1.  
 DR ProDom; PD000006; ABC transporter; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 SQ SEQUENCE 646 AA; 73170 MW; 3B5BE2D861F8DB2 CRC64;

Query Match 50.8%; Score 1703.5; DB 2; Length 646;  
 Best Local Similarity 53.4%; Pred. No. 4.5e-102;  
 Matches 350; Conservative 110; Mismatches 179; Indels 17; Gaps 8;

Qy 1 MSSNVEVPFIPVSGNTNGFPAATNDKAFTEGAVLSGHNIQYVKLKGFLPKRKYVE 60  
 Db 1 MSSNDPVIPIWIEROGDLPELTNTSDPETTLTKFAVLSFHNTSYRETVQSSP-PLRQOTR 59  
 Qy 61 -KEISNINIGIKKPGSLNIIPTGGKSSLDVLAAROPSGSLDVLINGAPRANFKC 119  
 Db 60 VMERISSISGIMPGSLNIIPTGGKSSLDVLAAROPSGSLDVLINGAPRANFKC 119  
 Qy 120 NSGVVQDDVWGLTVRENLOPSAALRLATMTNHEKERNINRYIOELGDKVADSKVG 179  
 Db 120 TSGYIPQNDVVLGTVVADNLEFSAALRLPMTVTDEKRRINEVLELHLEK----- 172  
 Qy 180 TQFIRGVSGGERKRTSIGMELITDPSIIFLDEPTTGLDSTANAVALLIKMSKQKRTII 239  
 Db 173 EQNVAPRKGLRKRTSIAMELTHEPIILFDPTTGLDRTTDTVISILRRMSMGKRTII 232  
 Qy 240 FSIHQPRISIFKLPTSLTLASGLMFGHQAQALGYEASAGHCEAVNPNADFLDIIIN 299  
 Db 233 FSIHQPRISIFKLPTSLTLASGLMFGHQAQALGYEASAGHCEAVNPNADFLDIIIN 292  
 Qy 300 GDSSTVALNREED-SEATEIIEPSKQDKPLEKLAIEYVNSFYETKALHQLSGGEKK 358  
 Db 293 GGF-----DTEBDGHADENEFERQHVQVQKWIICMSPRYHATRAARDQLLGEQKL 348  
 Qy 359 KKITVKEISYTSFCHQLRWVSKSPFNGLGNPQASIAQIIVTVGLVIGALVIGLKN 418  
 Db 349 DRSSAV-ETTCVTPFMHQIGWITRRSFKNGLGFPWTTTQVITIIIVILAVVGTARLLON 407  
 Qy 419 DSTGIONAGVLFITTNOCFSSVAVELFVVEKLEFHEHYSYGVYRVSIVYGLGLD 478  
 Db 408 VCTELQMRALILLGLGFCITSVTAGLEVIDQRFLEHHTSYGVYRVSIVYGLGLD 467  
 Qy 479 LPMTMLPSIIPTCIYVFMGLKPKADAFVVMFTLMMVAASASALAIAGQSVASVT 538  
 Db 468 IPRRLPPTIIFLLIYVFIAGLRISVRCGFTTFTIMLAISASSLSLGGAGENVAAITT 527

Qy 539 LMTICFVEMMI FSLGLVNLTTIASMLWQYFSIPRYGFTALQNEFLGONFCPLNAT 598  
 Db 528 LTVTVTVFVFMFLFSGSLSD-TCFLPVLWSIRKIFSIPHYGFPALLHNEFLGONFCBEYNT 586  
 Qy 599 GNNPC-NVATCTGEBYLVKQIDISPMGLMKHVALACMIYVFLTIAYLKLLFLK 653  
 Db 587 EVSRCHNVVICTGEBFLMIQIGIDISLWSGFWENHVALVCTMIITLITVOLLQVK 642

## RESULT 15

Q68HW7 PRELIMINARY; PRT; 1039 AA.  
 ID Q68HW7  
 AC Q68HW7;  
 DT 25-OCT-2004 (TREMBlrel. 28, Created)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE Similar to CA1379|CADP1 Candida albicans CADP1 ATP-dependent  
 permease.  
 GN ORFNames=DEHA0G113969;  
 OS Debaryomyces hansenii CBS767.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
 OX NCBI\_TaxId=284592;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS767;  
 RG Genolevures;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neugebauer C., Talla E.,  
 RA Goffard N., Prangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
 RA Barnay S., Branchet S., Beckerich J.M., Beyne E., Bleykaten C.,  
 RA Boiserame A., Boyer J., Catcolico U., Confantolero F., de Darvar A.,  
 RA Despons L., Fabre E., Fairhead C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicoud J.M., Nikolaki M., Ozias S., Oziar-Kalogeropoulos O.,  
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,  
 RA Swenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
 RA Zenou-Meyer C., Zivanovic I., Bolotin-Pukhara M., Thierly A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,  
 RA Wincker P., Souciet J.L.;  
 RT "Genome evolution in yeasts."  
 RL Nature 430:35-44(2004).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS767;  
 RA Genoscope;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to the ABC transporter family.

EMBL; CR382139; CAG90476.1; -  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:000524; F:ATP binding; IEA.  
 DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC transporter.  
 DR InterPro; IPR006209; EGF-like.  
 DR Pfam; PF00005; ABC tran. 1.  
 DR ProDom; PD000006; ABC transporter; 1.  
 DR SMART; SM00382; AAA\_1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_2.  
 DR ATP-binding; PS01186; EGF\_2; 2.  
 SQ SEQUENCE 1039 AA; 117062 MW; 51C861CAC3B092E4 CRC64;

Query Match 25.8%; Score 864; DB 2; Length 1039;  
 Best Local Similarity 31.8%; Pred. No. 2.4e-47;  
 Matches 219; Conservative 127; Mismatches 254; Indels 88; Gaps 13;

```
QY 27 DLKAFTEG---AVLSFHNICRYVKLSGFLPCRKPEKEILSNINGIMKP-GJNAILGPT 82
Db 372 ELNALNENFTPTLSFENISYEV--KSG-----QOVLDNVFGIVRPRECLAIWGS 420
QY 83 GGGKSSLLDLVLAARKDSGSGDLVING-APRPANFKNSGYVQDDVWVGTLTVRENLO 141
Db 421 GAOKTTLDDLKAGKKNKNGVNGQYINGNSLDPPDKYKMWGFVQDEQDLSTLTVETVL 480
QY 142 FSAALRLATMTNHEKNERINRVIOELDKVADSKVTOFIRGSGGERKRTSIGMELI 201
Db 481 NSALRLPRMSLRKRETRIVIEVLSERLISIKDRIGSDPKGICSGGERKRVSIACELV 540
QY 202 TDPSSILFLDEPTTGLDSSSTANAVLLIKRMSKQ-GRTIIFSIHOPRYSIFKLPDSLTLA 260
Db 541 TSPSILFLDEPTTGLDSSSTANAVVDCIVKLSRDEFTIIFTHIQPRSNIVSLFDKLLLS 600
QY 261 SGRLMFHGPQOELGYFESAGYHCEANNNPADPFLDI----- 297
Db 601 EGDLYSGDMIKCNDPFSKNGYKCPGLGYNLADYLDITVDHKLKVYKSKANNLLGSSN 660
QY 298 -----INGDSTAVALNREEDFKATEIIE-----PSKODK----- 326
Db 661 DLENLPGDADTDHQAFINNPATDIDITREMEHFAVHDEYNSTYHOKODKSGEETTYI 720
QY 327 PLIEKLAELIYVNSSFYKETAELHQLSGEKKKTIYFKEISYTTSFCHQLRWVSKRSFK 386
Db 721 QIRNKLPITYKESSLAELKQDINDLKNPRTIE--FKNQYORATFASQLLILASRTFK 777
QY 387 NLGNPQASIAQIIVVVLGVLGAIYFGKNDSTGIONRAGVLFPLTNQCFSSVSAVE 446
Db 778 NSYRNPKLLLSHYIISLAMGAFCGYLYDVENDISGFQNRGLFFVVLALFGSALTEH 837
QY 447 LFVVEKKLFHEIYISGYRVSSYFLGKLSDLLPMTLPSIIFTCIYVFMGLKPKRADAF 506
Db 838 SFSTERIIFIRERANNYYHPLSYLSKIIDIIPLRVLPRIILISITYPVLGTLMENNGF 897
QY 507 FVMMFTLMVAVYASSMALAIAAGQSVSVATLMTICFVPMIIFSGLLVNLTTIASWLS 566
Db 898 LKTIILVLFENISIAAEVLVVGILKEPGTSTWGVLLFSLFAGLFINSENVAVQIK 957
QY 567 WLOYPISIPRYGFALQHNE--FLGQNFPCGLNATGNNPCNVATCTGEETVLVQGDIDSP 623
Db 958 WLEWVSIFHYAYBALAINEVKDLILREKKYGLS-----IEVPAVILSTFGFVNSA 1008
QY 624 WGLMKNHVALACMIVIFLTIAYLKLJFL 651
Db 1009 Y--WKDVSCIAASTIAFLILIGYIFLHWL 1034
```

Search completed: June 6, 2005, 09:43:41  
Job time : 178 secs